

# Pre- and post-selection, weak values, and contextuality

Jeff Tollaksen, Department of Physics and  
 Department of Computational and Data Sciences,  
 College of Science, George Mason University, Fairfax, VA 22030<sup>1</sup>

## Abstract

By analyzing the concept of contextuality (Bell-Kochen-Specker) in terms of pre-and-post-selection (PPS), it is possible to assign definite values to observables in a new and surprising way. Physical reasons are presented for restrictions on these assignments. When measurements are performed which do not disturb the pre- and post-selection (i.e. weak measurements), then novel *experimental* aspects of contextuality can be demonstrated including a proof that every PPS-paradox with definite predictions implies contextuality. Certain results of these measurements (eccentric weak values with e.g. negative values outside the spectrum), however, cannot be explained by a “classical-like” hidden variable theory.

## Contents

<b>1</b>	<b>INTRODUCTION</b>	<b>2</b>
<b>2</b>	<b>PPS, CONTEXTUALITY, AND THE 3-BOX-PARADOX</b>	<b>4</b>
2.1	Weak Measurements . . . . .	5
2.2	Contextuality and PPS-paradoxes . . . . .	7
2.2.1	HVTs and PPSs: . . . . .	8
2.2.2	Logical-PPS-paradoxes imply contextuality through WMs: . . . . .	9
<b>3</b>	<b>PPS AND CONTEXTUALITY IN 4-D</b>	<b>10</b>
3.1	Review of Mermin’s 4-D BKS theorem . . . . .	10
3.2	ABL, VAA and BKS nonets . . . . .	12
3.2.1	Ascertaining the results of any one of the 9 observables: . . . . .	12
3.2.2	Ascertaining the results of <i>products</i> of the 9 observables: . . . . .	13
3.2.3	Ascertaining the results of any one of 16 observables through the generalized . . . . .	13
3.2.4	A physical reason for restrictions on these assignments . . . . .	16
3.3	Non-classical Weak Values for the 4D BKS nonets . . . . .	17
3.4	Non-classical weak values for EPR and Peres/BKS . . . . .	20
<b>4</b>	<b>PPS AND CONTEXTUALITY IN HIGHER DIMENSIONS</b>	<b>23</b>
4.1	GHZ and PPS . . . . .	23
4.2	Weak Values in GHZ state . . . . .	24
<b>5</b>	<b>CONCLUSION</b>	<b>25</b>

---

<sup>1</sup>email: jtollaks@gmu.edu

# 1 INTRODUCTION

A traditional concept of the quantum state  $|\Psi_{\text{in}}\rangle$  is that it generally provides only statistical information about the outcome of an Ideal Measurement (IM). Therefore many authors have proposed that the quantum state could be “completed” by a hidden-variable-theory (HVT). A natural category of HVT assigns definite values to all possible observables of a system at all times has a similar relationship to Quantum Mechanics (QM) as Classical Mechanics (CM) has to Classical Statistical Mechanics (CSM). I.e. QM can be understood in terms of a deeper theory, the HVT. The relationship between CM and CSM is relatively simple because an ideal classical measurement precisely measures a property of a system, *without* affecting the system under study. Measurement of one property will not interfere with measurement of another property (i.e. measurement does not depend on context) and the state of the system can then be said to exist when we are not observing it. There is thus a simple relation between the theory and underlying physical processes (the outcome of measurements directly tells us what values to assign to all variables of the theory).

Moving over to QM, there are two general constraints on any HVT which reproduces QM: a) the Bell-Kochen-Specker theorem (BKS) [21, 24] and Gleason’s [33] theorem showed that any HVT must be contextual and b) Bell’s theorem [20], showed that any HVT must be non-local. Gleason and BKS proved that one cannot assign unique answers to yes-no questions (posed to single systems) in such a way that one can think that measurement simply reveals the answer as a pre-existing property that was intrinsic solely to the quantum system itself. Some versions of BKS depend only on the structure of observables, while some versions also rely on the state and we therefore define a “value function”  $V_{\vec{\psi}}(\hat{A})$  (the specification of the HVT) as the assignment of a value to an observable  $\hat{A}$  when an individual system is in the state  $\vec{\psi}$ . BKS assumed that  $V_{\vec{\psi}}(\hat{A})$  should satisfy:

$$V_{\vec{\psi}}(F\{\hat{A}\}) = F\{V_{\vec{\psi}}(\hat{A})\} \quad (1.1)$$

That is, any functional relation of an operator that is a member of a commuting subset of observables must also be satisfied if one substitutes the values for the observables into the functional relations. E.g. if a system is characterized by commuting observables  $\hat{A}_1$  and  $\hat{A}_2$  then condition 1.1 requires that all the *relationships or functions* between these operators should also be satisfied when  $V_{\vec{\psi}}(\hat{A}_1)$  and  $V_{\vec{\psi}}(\hat{A}_2)$  are substituted into the same functional relations. This condition determines the sum and product rules:

$$V_{\vec{\psi}}(\hat{A}_1 + \hat{A}_2) = V_{\vec{\psi}}(\hat{A}_1) + V_{\vec{\psi}}(\hat{A}_2) \quad (1.2)$$

$$V_{\vec{\psi}}(\hat{A}_1 \hat{A}_2) = V_{\vec{\psi}}(\hat{A}_1) V_{\vec{\psi}}(\hat{A}_2) \quad (1.3)$$

HVTs which meet these conditions are **noncontextual**: all yes-no questions can be associated with a value assignment  $V_{\vec{\psi}}$  which provides a single unique answer, irrespective of the set of other commuting yes-no questions that it is associated with. BKS showed that attempting such an assignment to some observables is inconsistent: in any system (of dimension greater than 2) the  $2^n$  possible “yes-no” assignments (to the  $n$  projection operators representing the yes-no questions) cannot be compatible

with the sum 1.2 and product 1.3 rules for all orthogonal resolutions of the identity. E.g., consider a complete set of spectral projectors of an operator  $\hat{A}$  with discrete eigenvalues, so that  $\hat{P}_i = \hat{P}_{A=a_i}$ , such that  $\sum_{i=1}^n V_{\vec{\psi}}(\hat{P}_i) = 1$ . Then only one of the projection operators can give a “yes” assignment ( $\hat{P}_i = 1$ ) and the rest have to be “no” (i.e.  $\hat{P}_i = 0$ ). However,  $\vec{\psi}$  can be decomposed into many different basis sets, and the value that  $V_{\vec{\psi}}$  assigns must be independent of the particular basis. BKS showed that this cannot be done.

The principal result of this article is to question whether BKS is just a formal result (i.e. negative statements concerning the impossibility of a classical-like “non-contextual-HVTs”) or if BKS has new positive aspects including experimental consequences. We probe this by utilizing the natural connection between counterfactual statements and pre-and-post-selection (PPS). We then connect contextuality with issues that can be probed experimentally by weak measurements. PPSs were originally probed with the time-symmetric re-formulation of Quantum Mechanics (TSQM, introduced by Aharonov, Bergmann and Lebowitz a/k/a ABL [2]). To be useful and interesting, any re-formulation of QM should meet several criteria such as those met by TSQM:

- TSQM is consistent with all the predictions made by standard QM,
- TSQM brings out features in QM that were missed before: e.g., ABL considered measurement situations *between* two successive IMs in which the transition from a state  $|\Psi_{\text{in}}\rangle$  (pre-selected at a time  $t_{\text{in}}$ ) to a state  $|\Psi_{\text{fin}}\rangle$  (post-selected at a later time  $t_{\text{fin}}$ ) is generally disturbed by an intermediate precise measurement. Post-selection reflects a unique aspect of QM in that measurement results are not determined by equations of motion and initial conditions. A subsequent theoretical development arising out of the ABL work was the introduction of the “Weak Value” (WV) of an observable which was probed by a new type of quantum measurement called the “Weak Measurement” (WM) [5]. WM experiments have been performed and results are in very good agreement with theoretical predictions,
- TSQM lead to simplifications in calculations (as occurred with the Feynman re-formulation) and stimulated discoveries in other fields: e.g. ABL influenced work in cosmology (e.g. Gell-Mann and Hartle [23]); in superluminal tunneling (Chiao [22] and Steinberg [34]); in quantum information (e.g. the quantum random walk [8] or cryptography [16, 15]), etc.
- TSQM suggests generalizations of QM that were missed before - e.g. a new solution [32] to the quantum measurement problem.

Using TSQM, we show how to assign definite values to sets of BKS observables in a new and surprising way. We also show how measurement disturbance can arise in a new way when value assignments depend on both the pre- and post-selection. An “intriguing” physical reason is presented to explain why this scheme cannot be applied to 2 or more IMs: the 2 IMs interfere with each other because some assignment of eigenvalues to operators are based on just one of the two vectors (i.e. either the pre- *or* the post-selected vector) while some assignment of eigenvalues are based on

both vectors (i.e. both the pre-selected *and* the post-selected vectors, what we call diagonal-PPS). In addition, we show that when measurements are performed which do not disturb the PPS (WMs), then novel *experimental* aspects of contextuality can be demonstrated, including a proof that every PPS-paradox with definite predictions implies contextuality. We also demonstrate an isomorphism between WV<sub>s</sub> in BKS situations and EPR entanglement. Certain results of WMs (eccentric WV<sub>s</sub> with outside the eigenvalue spectrum), however, cannot be explained by a “classical-like” HVT.

## 2 PPS, CONTEXTUALITY, AND THE 3-BOX-PARADOX

The 3-box-PPS-paradox ([3], verified experimentally [35]) uses a *single* quantum particle that is placed in a superposition of being in 3 closed, separated boxes. The particle is pre-selected in  $|\Psi_{\text{in}}\rangle = 1/\sqrt{3} (|A\rangle + |B\rangle + |C\rangle)$ , where  $|A\rangle$ ,  $|B\rangle$  and  $|C\rangle$  denote the particle localized in boxes  $A$ ,  $B$ , or  $C$ , respectively. The particle is post-selected in the state  $|\Psi_{\text{fin}}\rangle = 1/\sqrt{3} (|A\rangle + |B\rangle - |C\rangle)$ . If an ideal (i.e. von Neumann) measurement is performed on box  $A$  in the intermediate time (e.g. we open the box), then the probability to find the particle in box  $A$  is 1, i.e.  $\hat{\mathbf{P}}_A = |A\rangle\langle A| = 1$ , given by ABL [2]<sup>2</sup>:

$$\text{Prob}(a_j, t|\Psi_{\text{in}}, t_{\text{in}}; \Psi_{\text{fin}}, t_{\text{fin}}) = \frac{|\langle\Psi_{\text{fin}}, t|a_j\rangle\langle a_j|\Psi_{\text{in}}, t\rangle|^2}{\sum_{a'} |\langle\Psi_{\text{fin}}, t|a'\rangle\langle a'|\Psi_{\text{in}}, t\rangle|^2} \quad (2.4)$$

which in this case yields,  $\text{Prob}(\hat{\mathbf{P}}_A) = \frac{|\langle\Psi_{\text{fin}}|\hat{\mathbf{P}}_A|\Psi_{\text{in}}\rangle|^2}{|\langle\Psi_{\text{fin}}|\hat{\mathbf{P}}_A|\Psi_{\text{in}}\rangle|^2 + |\langle\Psi_{\text{fin}}|\hat{\mathbf{P}}_B + \hat{\mathbf{P}}_C|\Psi_{\text{in}}\rangle|^2} = 1$ . This can also be seen intuitively without using ABL: if the particle is not found in box  $A$ , then the initial state  $|\Psi_{\text{in}}\rangle$  would be projected onto  $1/\sqrt{2} (|B\rangle + |C\rangle)$ , but this is orthogonal to the final state  $|\Psi_{\text{fin}}\rangle$ . Therefore the particle must be found in box  $A$ . Similarly, the probability to find the particle in box  $B$  is 1, i.e.  $\text{Prob}(\hat{\mathbf{P}}_B = 1) = 1$ , and this is the essence of the “paradox”.

Now  $\text{Prob}(\hat{\mathbf{P}}_A = 1) = 1$  if only box  $A$  is opened, while  $\text{Prob}(\hat{\mathbf{P}}_B = 1) = 1$  if only box  $B$  is opened. When we measure both box  $A$  and box  $B$ , then the particle will not be found in both boxes, i.e.  $\hat{\mathbf{P}}_A\hat{\mathbf{P}}_B = 0$ . But  $\hat{\mathbf{P}}_A$  and  $\hat{\mathbf{P}}_B$  commute with each other, so how is it possible that measurement of one box can disturb the measurement of another? The reason suggested here is that in order to deduce  $\hat{\mathbf{P}}_A = 1$  from ABL (or  $\hat{\mathbf{P}}_B = 1$ ), we used information from both the pre-and the post-selected vectors (a situation we call diagonal-PPS); when we actually measure  $\hat{\mathbf{P}}_A$ , then this measurement will limit the “propagation” of the 2-vectors that were relied on to ascertain intermediate values (see fig. 1.b). If we then subsequently were to measure  $\hat{\mathbf{P}}_B$ , the necessary information from both the pre- and post-selected vectors is no longer available (i.e. information from  $t_{\text{in}}$  cannot propagate beyond the measurement

---

<sup>2</sup>The time-symmetry in ABL can already be seen from the Born formula  $\text{Prob}(a_j, t|\Psi_{\text{in}}, t_{\text{in}}) = |\langle a_j|U(t, t_{\text{in}})|\Psi_{\text{in}}\rangle|^2$ . Instead of  $|\langle a_1| \exp(-iH\Delta t)\Psi_{\text{in}}\rangle|^2$  for a particular outcome  $a_1$ , one may equivalently say that the probability is  $|\langle a_1 \exp(+iH\Delta t) |\Psi_{\text{in}}\rangle|^2$ , i.e. that one applies the time evolution operator to evolve  $a_1$  from  $t$  to  $t_{\text{in}}$ , which is the time reverse of the first picture.

of  $\hat{\mathbf{P}}_A$  at time  $t_1$  due to the disturbance caused by an IM). Thus, even though  $\hat{\mathbf{P}}_A$  and  $\hat{\mathbf{P}}_B$  commute, measurement of one can disturb measurement of the other with diagonal-PPSs, and this is related to a violation of the product rule 1.3. (In general, if  $|\Psi_1\rangle$  is an eigenvector of  $\hat{A}$  with eigenvalue  $a$  and  $|\Psi_2\rangle$  is an eigenvector of  $\hat{B}$  with eigenvalue  $b$  and  $[\hat{A}, \hat{B}] = 0$ , then if  $\hat{A}$  and  $\hat{B}$  are known only by either pre-selection or post-selection, then the product rule is valid, i.e.  $\hat{A}\hat{B} = ab$ . However if  $\hat{A}$  and  $\hat{B}$  are known by both pre-selection and post-selection, then the product rule is not valid, i.e.  $\hat{A}\hat{B} \neq ab$ , i.e. they can still disturb each other, even though they commute. [29])

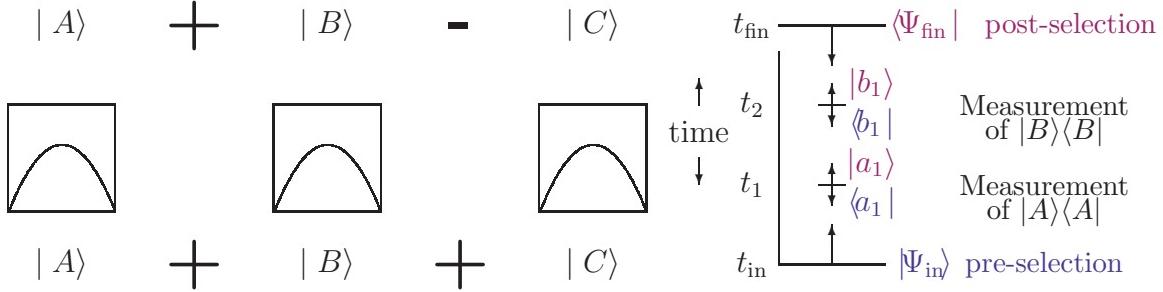


Figure 1: a) pre-selected vector  $|\Psi_{\text{in}}\rangle = 1/\sqrt{3} (|A\rangle + |B\rangle + |C\rangle)$  propagates forwards in time from  $t_{\text{in}}$  to  $t_1$ , and post-selected vector  $|\Psi_{\text{fin}}\rangle = 1/\sqrt{3} (|A\rangle + |B\rangle - |C\rangle)$  propagating backwards in time from  $t_{\text{fin}}$  to  $t_2$ . b) IM of  $\hat{\mathbf{P}}_A$  at  $t_1$  and of  $\hat{\mathbf{P}}_B$  at  $t_2$ .

## 2.1 Weak Measurements

WMs can be quantified in the quantum measurement theory developed by von Neumann [38]: to measure an observable  $\hat{A}$  of the system, one may use an interaction Hamiltonian of the form  $H_{\text{int}} = -\lambda(t)\hat{Q}\hat{A}$  where  $\hat{Q}$  is an observable of the measuring device (MD) and  $\lambda(t)$  is a coupling constant which is non-zero only during a short time  $(0, T)$ . Using the Heisenberg equations of motion for the momentum  $\hat{P}$  of MD (conjugate to the position  $\hat{Q}$ ), we see that  $\hat{P}$  changes according to  $\frac{d\hat{P}}{dt} = \lambda(t)\hat{A}$ . Integrating this, we see that  $P(T) - P(0) = \lambda A$  where  $\int_0^T \lambda(t)dt = \lambda$ . To make a more precise determination of  $\hat{A}$  requires that either a)  $P(0)$  and  $P(T)$  are more precisely defined or b)  $\lambda$  is large.

A WM can be characterized by either a)  $\hat{P}$  of MD is measured to a finite precision  $\Delta P$ , (which limits the disturbance by a finite amount  $\Delta Q \geq 1/\Delta P$ ) or b) small  $\lambda$ . After the WM interaction, the system is post-selected. In this regime, the measurement becomes less precise because the uncertainty  $\Delta P$  in the position of the pointer is larger than the difference in the shifts of the pointer  $\lambda a_i$  corresponding to the different eigenvalues and thus the shift in MD is much smaller than its uncertainty. The simplest derivation of the WV result is with the second approach, i.e.  $\lambda$  small ( $\int \lambda(t)dt = \lambda \ll 1$ ):

$$\langle \Psi_{\text{fin}} | e^{-i\lambda\hat{Q}\hat{A}} | \Psi_{\text{in}} \rangle \approx \langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle \{1 + i\lambda\hat{Q}A_w\} \approx \langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle e^{-i\lambda\hat{Q}A_w} \quad (2.5)$$

The final state of MD is almost unentangled with the system and is shifted by the WV  $A_w$  (assuming without loss of generality that the state of the MD is a Gaussian

with spreads  $\Delta \equiv \Delta P = \Delta Q = 1$ ):

$$\tilde{\Phi}_{\text{fin}}^{\text{MD}}(P) \rightarrow \langle \Psi_{\text{fin}} | e^{-i\lambda \hat{Q}\hat{A}} | \Psi_{\text{in}} \rangle \tilde{\Phi}_{\text{in}}^{\text{MD}}(P) = \exp \left\{ -(P - \lambda A_w)^2 \right\} \quad (2.6)$$

$$\text{where } A_w = \frac{\langle \Psi_{\text{fin}} | \hat{A} | \Psi_{\text{in}} \rangle}{\langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle} \quad (2.7)$$

For the PPS cases considered, we do not need to perform the above calculation to obtain the WV. Instead, we can easily ascertain the WVs due to the following theorems:

**Theorem 1:** The sum of the WVs is equal to the WV of the sum:

$$\text{if } (\hat{\mathbf{P}}_A)_w = (\hat{\mathbf{P}}_B + \hat{\mathbf{P}}_C)_w \text{ then } (\hat{\mathbf{P}}_A)_w = (\hat{\mathbf{P}}_B)_w + (\hat{\mathbf{P}}_C)_w \quad (2.8)$$

Proof: this follows simply from the linearity of the operators

$$\frac{\langle \Psi_{\text{fin}} | \hat{\mathbf{P}}_B + \hat{\mathbf{P}}_C | \Psi_{\text{in}} \rangle}{\langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle} = \frac{\langle \Psi_{\text{fin}} | \hat{\mathbf{P}}_B | \Psi_{\text{in}} \rangle}{\langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle} + \frac{\langle \Psi_{\text{fin}} | \hat{\mathbf{P}}_C | \Psi_{\text{in}} \rangle}{\langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle} \quad (2.9)$$

**Theorem 2:** If a single IM of an observable  $\hat{\mathbf{P}}_A$  is performed between the pre- and the post-selection, then if the outcome is definite (e.g.  $\text{Prob}(\hat{\mathbf{P}}_A = 1) = 1$ ) then the WV is equal to this eigenvalue (e.g.  $(\hat{\mathbf{P}}_A)_w = 1$ ) [6].

Proof: Given that  $\hat{\mathbf{P}}_A = \sum_n a_n |\alpha_n\rangle\langle\alpha_n|$ , if an eigenvalue, e.g.  $\hat{\mathbf{P}}_A = a_n$ , is obtained with certainty, then for  $n \neq m$ ,  $\hat{\mathbf{P}}_A \equiv |\alpha_m\rangle\langle\alpha_m| = 0$  because the probability to obtain another eigenvalue by ABL is  $\propto \langle \Psi_{\text{fin}} | \alpha_m \rangle \langle \alpha_m | \Psi_{\text{in}} \rangle = 0$ . In this case, the WV  $(\hat{\mathbf{P}}_A)_w = (|\alpha_m\rangle\langle\alpha_m|)_w = \frac{\langle \Psi_{\text{fin}} | \alpha_m \rangle \langle \alpha_m | \Psi_{\text{in}} \rangle}{\langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle} = 0$ . In addition,

$$\sum_m \frac{\langle \Psi_{\text{fin}} | \alpha_m \rangle \langle \alpha_m | \Psi_{\text{in}} \rangle}{\langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle} = 1 \quad (2.10)$$

because  $\sum_m |\alpha_m\rangle\langle\alpha_m| = 1$ . But since  $\langle \Psi_{\text{fin}} | \alpha_m \rangle \langle \alpha_m | \Psi_{\text{in}} \rangle = 0$  for  $n \neq m$ , the only term left is  $n$ . Therefore, the WV is 1, the same as the ideal value.

From Theorem 2, we know the following WVs in the 3-box-paradox with certainty:

$$(\hat{\mathbf{P}}_A)_w = 1, \quad (\hat{\mathbf{P}}_B)_w = 1, \quad \hat{\mathbf{P}}_{\text{total}} = (\hat{\mathbf{P}}_A + \hat{\mathbf{P}}_B + \hat{\mathbf{P}}_C)_w = 1. \quad (2.11)$$

Using theorem 1, we obtain:

$$\begin{aligned} (\hat{\mathbf{P}}_C)_w &= \frac{\langle \Psi_{\text{fin}} | \hat{\mathbf{P}}_{\text{total}} - \hat{\mathbf{P}}_A - \hat{\mathbf{P}}_B | \Psi_{\text{in}} \rangle}{\langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle} \\ &= (\hat{\mathbf{P}}_A + \hat{\mathbf{P}}_B + \hat{\mathbf{P}}_C)_w - (\hat{\mathbf{P}}_A)_w - (\hat{\mathbf{P}}_B)_w = -1. \end{aligned} \quad (2.12)$$

This surprising theoretical prediction of TSQM has been verified experimentally using photons [35]. What interpretation should be given to  $(\hat{\mathbf{P}}_C)_w = -1$ ? One may speculate for formal reasons that this corresponds to a “negative probability.” However, as will be shown subsequently, this interpretation cannot have any experimental

meaning. On the other hand, we can give it a different interpretation that **does** have an experimental meaning if we perform any WM which is sensitive to the projection operator  $\hat{\mathbf{P}}_C$ . In this case, we will observe the opposite effect from those cases in which the projection operator is positive. This suggests that there is  $-1$  particle in box  $C$ , e.g. a WM of the amount of charge in box  $C$  in the intermediate time will yield a negative charge (assuming it is a positively charged particle).

## 2.2 Contextuality and PPS-paradoxes

In [30] it was first pointed out and extensively discussed and later proven by Leifer and Spekkens [26], that whenever there is a logical-PPS-paradox (as in the 3-box-paradox), then there is a related proof of contextuality. The proof considers “all the measurements defined by the PPS-paradox –the preselection, the post-selection, and the alternative possible intermediate measurements—as alternative possible measurements at a single time.” They show a direct connection between each of the 8 vectors in the 3-box-paradox and the 8 vectors in the Clifford/BKS-proof: it is readily seen ([26]’s figure) that no HVT assignment of 0’s (black circles) and 1’s (white circles) can be made that is consistent with the orthogonality relations, since in order for non-contextuality to hold, no 2 orthogonal pairs can be white, which is violated by the central 2 circles.

On the one hand, we will show how definite values can be assigned to observables (verified by IMs on PPSs) in surprising ways. While this assignment suggests novel connections between what could be said about the state before the IM and after, in general, the IM creates a disturbance and thus creates an uncertain relationship between the state before and after, reflecting Bub and Brown’s [47] concern “ensembles which have been preselected and post-selected via an arbitrary intervening measurement...are not well defined without specification of the intervening measurement.” However, this is not the case for WMs (PPS ensembles are well-defined for any intermediate WM). In addition, we show [30] how measurement disturbance can arise in new ways for IMs of commuting observables in PPS situations, and so it was also argued [30, 25] that a noncontextual-HVT can reproduce QM if we allow for a disturbance of the HVs: “the possibility of measurement disturbance blocks the conclusion that a PPS-paradox is itself a proof of the contextuality of HVTs.” The motivation [25] behind this assertion was the belief that PPS-paradoxes could be explained entirely within CM [40, 41] and that contextuality should not be regarded as fundamental in a classical picture of reality. Nevertheless, the “paradoxical,” i.e. non-classical, nature of the 3-box-paradox was recently re-affirmed in terms of IMs [44]. In this article, we expand this point mainly by showing how non-classical WVs can be empirically demonstrated in PPS-paradoxes. E.g. WMs on box  $C$  will record the “paradoxical,” non-classical outcome of  $-1$ . WVs like this outside the eigenvalue spectrum cannot be reproduced from any positive definite probability distribution of eigenvalues [31, 14, 30].

Finally, with WMs, there is no measurement disturbance, yet non-classical results are still obtained. In §2.2.2 we show how WMs allow us to make a stronger connection between PPSs and contextuality. This weakens the assertion that [25]

“PPS paradoxes do not require contextuality for their explanation but do require measurement-disturbance.”

### 2.2.1 HVTs and PPSs:

In [25], quantum states are given as probability distributions  $\mu$  over HVTs  $\lambda$ , such that  $\int_{\Omega} \mu(\lambda) d\lambda = 1$ , where  $\Omega$  is the set of possible HVTs and measurements are characterized by idempotent indicator functions  $\chi_j^M : \Omega \rightarrow \{0, 1\}$ , such that  $\sum_j \chi_j^M(\lambda) = 1$ . The probability to obtain an outcome  $j$  for a random variable  $X$  when a measurement  $M$  is performed on a system in  $\mu$  then is

$$p_{\mu}(X_M = j) = \int_{\Omega} \chi_j^M(\lambda) \mu(\lambda) d\lambda \quad (2.13)$$

In addition, a transition from  $\omega$  to some other HVT state  $\lambda$  could result from the measurement process [25], and this is modeled by a transition probability  $D_j^M(\lambda, \omega)$  such that  $\int_{\Omega} D_j^M(\lambda, \omega) d\lambda = 1$ . This approach to HVTs may then be applied [26] to ABL:

$$p_{\text{HVT}}(X_M = k | A_{\text{pre}}, A_{\text{post}}, M) = \frac{\int_{\Omega} \chi_{\text{post}}(\lambda) \Gamma_k^M(\lambda, \omega) \mu_{\text{pre}}(\omega) d\omega d\lambda}{\int_{\Omega} \chi_{\text{post}}(\lambda) (\Gamma_k^M(\lambda, \omega) + \Gamma_{-k}^M(\lambda, \omega)) \mu_{\text{pre}}(\omega) d\omega d\lambda} \quad (2.14)$$

WMs and WVs can also be described in this language (the usual projective measurement typically utilized in quantum experiments are a special case of WMs [48]) and this will be developed in a subsequent article [46]. Some general considerations on possible relationships between HVTs and WVVs are:

- **Epistemic nature of probability:** If the probabilistic nature of  $D_j^M(\lambda, \omega)$  and  $\chi_j^M(\lambda)$  are of epistemic origin, then this HVT approach [25] requires that non-commuting observables (such as  $p$  and  $x$ ) can have a “simultaneous” precise reality, as suggested, e.g. by the Wigner-Moyal method. If we require that any theoretical formalism should include exactly what can be measured (no more and no less), then it should be possible to make measurements on these projections. While such densities do give the correct average of a function, i.e.  $\int \rho(x, p) f(x, p) dx dp$  (thus appear to behave as proper densities), they also have un-physical aspects, i.e. mathematical artefacts, when the densities become negative. The reason (as will be shown subsequently [46]) is that if we attempt to actually measure such “negative” properties, then the result does not correspond to a physical observable in Hilbert Space. E.g. if we did try to project on  $p$  and  $x$  as densities simultaneously, then we obtain the parity operator, taking a generic  $\psi(x)$  to  $\psi(-x)$ . To see this, we translate the classical projection  $p = 0$  and  $x = 0$  into QM:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\alpha x + i\beta p} d\alpha d\beta \underset{QM}{\Rightarrow} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\frac{i\alpha\beta}{2}} e^{i\alpha\hat{x}} e^{i\beta\hat{p}} d\alpha d\beta \quad (2.15)$$

Consider applying this to a generic wavefunction. First, the exponential,  $e^{i\beta\hat{p}}$ ,

translates  $\psi(x)$ . Integrating then over  $\alpha$  produces a delta function:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\frac{\alpha\beta}{2}} e^{i\alpha x} \psi(x + \beta) d\alpha d\beta = \int_{-\infty}^{\infty} \left\{ \underbrace{\int_{-\infty}^{\infty} e^{i\alpha(x+\frac{\beta}{2})} d\alpha}_{\delta(x+\frac{\beta}{2})} \right\} \psi(x + \beta) d\beta \quad (2.16)$$

Finally, integrating over  $\beta$ , we obtain  $\beta = -2x$ , and thus  $\psi(x - 2x) = \psi(-x)$ . Therefore, the quantum analog of the classical projection does not correspond to a quantum projector: it corresponds to a highly non-local result, the parity operator.

- **Ontological nature of probability:** In contrast to this non-physical aspect, non-classical WVs (e.g. negative WVs for projection operators) **can** be seen experimentally. There are additional advantages besides this greater harmony between measurements and theory [46]. E.g. if the quantum/probabilistic nature of  $D_j^M(\lambda, \omega)$  is taken to be fundamental, then a natural/axiomatic explanation can be obtained [46] for 2 seemingly opposite situations: a) the contextual dependence of value assignments on details of MD and b) the fact that the statistics do not depend on the details of MD. These can be harmonized as a consistency check with causality [46] as indicated by inserting a complete set of states  $\{|\Psi_{\text{fin}}\rangle_j\}$  into  $\bar{A}$ :

$$\bar{A} = \langle \Psi_{\text{in}} | \left[ \sum_j |\Psi_{\text{fin}}\rangle_j \langle \Psi_{\text{fin}}|_j \right] \hat{A} | \Psi_{\text{in}} \rangle = \sum_j |\langle \Psi_{\text{fin}}|_j \Psi_{\text{in}} \rangle|^2 \frac{\langle \Psi_{\text{fin}}|_j \hat{A} | \Psi_{\text{in}} \rangle}{\langle \Psi_{\text{fin}}|_j \Psi_{\text{in}} \rangle} \quad (2.17)$$

Thus, one can think of  $\bar{A}$  for the whole ensemble as being built out of pre- and post-selected states in which the WV is multiplied by a probability for post-selection. The fluctuation in the system then is also relevant for the probability to obtain different post-selections: as the fluctuation in the system increases, the probability of a rare or eccentric post-selection also increases. However an attempt to see this through WMs will require the spread in the MD to be increased and this increases the probability of seeing the strange result as an error of the MD.

- **Time-symmetry versus time-asymmetry:** The HVT approach utilized in [25] focuses on the probability of post-selection and brings in an element of time-asymmetry, apparently endemic to a Bayesian approach [14].

### 2.2.2 Logical-PPS-paradoxes imply contextuality through WMs:

A logical-PPS-paradox occurs if  $p_{\text{HVT}} = 1$  for several incompatible situations. This can occur if

$$\int_{\Omega} \chi_{\text{post}}(\lambda) \Gamma_{\neg k}^M(\lambda, \omega) \mu_{\text{pre}}(\omega) d\omega d\lambda = 0 \quad (2.18)$$

However, Leifer and Spekkens [25] argue that the elements in the PPS-paradox have nothing to do with contextuality because their interpretation is different: “..we did not show that a logical PPS paradox is *itself* a proof of contextuality; measurements that are temporal successors in the PPS paradox must be treated as counterfactual alternatives in the proof of contextuality. This distinction is critical, since an earlier measurement can cause a disturbance to the ontic state that is monitored by a later measurement”. However, WMs (instead of IMs) can give an empirical meaning to “alternative possible measurements at a single time”. Although these alternatives are usually regarded as counter-factual, they will all be true simultaneously with WMs.

**Theorem:** Logical-PPS-paradoxes imply contextuality through WMs. Proof: Theorem 2 allows us to state that all counterfactual statements which maintain the occurrence of an outcome with certainty will all be true simultaneously when they are measured weakly. Theorem 2 is applicable to the precise elements utilized in the contextuality proof [26]. In addition, given that WV are by definition independent of the type of WM and given eq. 2.17, and since WV can violate the algebraic conditions (the product rule) required for BKS and noncontextual-HVTs [25], we have now proven that logical-PPS-paradoxes which assign definite probabilities (of 0 or 1) via ABL, are in fact proofs of contextuality if all the “alternative possible intermediate measurements” are performed weakly and the argument that their interpretation is different does not apply.

We have thereby mitigated the attempt to explain these PPS “paradoxes” as a result of disturbance (since there is no disturbance with WMs) and have strengthened [26] the connection between these “paradoxes” (which we will argue cannot be reproduced by a “classical-like” HVT) and contextuality.

### 3 PPS AND CONTEXTUALITY IN 4-D

In the 3-box-paradox, the product of observables was always definite, i.e.  $\hat{\mathbf{P}}_A \hat{\mathbf{P}}_B = 0$  and the proof of contextuality was state dependent. In this section, we consider a slightly different situation (4D BKS nonets) in which the product of observables can give 2 different values. Except for this difference, this 4D BKS nonet example is similar to the 3-box-paradox in that we shall also analyze them in terms of PPSs thereby revealing surprising predictions for IMs and WMs and will demonstrate the identical issues of diagonal-PPS measurements, violation of the product rule, contextuality, and WMs which cannot be explained by a noncontextual-HVT.

#### 3.1 Review of Mermin’s 4-D BKS theorem

We consider Mermin’s version of BKS with a set of 9 observables. It is intuitive [28] to represent all the “functional relationships between mutually commuting subsets of the observables,” i.e.  $V_{\vec{\psi}}(F\{\hat{A}\}) = F\{V_{\vec{\psi}}(\hat{A})\}$ , by drawing them in fig. 2 and arranging them so that all the observables in each row (and column) commute with all the other observables in the same row (or column).

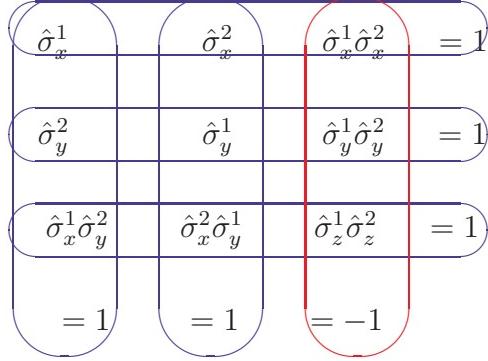


Figure 2: 4-D BKS example

Individually, each of the 9 observables depicted in fig. 2 has eigenvalues  $\pm 1$ . A noncontextual-HVT requires an assignment of values that is  $\pm 1$ . In addition, eq. 1.1 requires that the value assigned to the product of all three observables in any row or column must obey the same identities that the observables themselves satisfy, i.e. the product of the values assigned to the observables in each oval yields a result of  $+1$  except in the last column which gives  $-1$ . (The value assignments are given by  $V_{\vec{\psi}}(\hat{\sigma}_x^1) = \langle \hat{\sigma}_x^1 \otimes I^2 \rangle$ ,  $V_{\vec{\psi}}(\hat{\sigma}_x^2) = \langle I^1 \otimes \hat{\sigma}_x^2 \rangle \dots V_{\vec{\psi}}(\hat{\sigma}_z^1) = \langle \hat{\sigma}_z^1 \otimes \hat{\sigma}_z^2 \rangle$ ). Computing column 3 of fig. 2:

$$\begin{aligned} \{\hat{\sigma}_x^1 \hat{\sigma}_x^2\} \{\hat{\sigma}_y^1 \hat{\sigma}_y^2\} \{\hat{\sigma}_z^1 \hat{\sigma}_z^2\} &= \hat{\sigma}_x^1 \underbrace{\hat{\sigma}_x^2 \hat{\sigma}_y^1}_{\text{commute } so \leftrightarrow} \hat{\sigma}_y^2 \hat{\sigma}_z^1 \hat{\sigma}_z^2 = \underbrace{\hat{\sigma}_x^1 \hat{\sigma}_y^1}_{=i\hat{\sigma}_z^1} \underbrace{\hat{\sigma}_x^2 \hat{\sigma}_y^2}_{=i\hat{\sigma}_z^2} \hat{\sigma}_z^1 \hat{\sigma}_z^2 \\ &= i\hat{\sigma}_z^1 \underbrace{i\hat{\sigma}_z^2 \hat{\sigma}_z^1}_{\text{commute } so \leftrightarrow} \hat{\sigma}_z^2 = i\hat{\sigma}_z^1 i\hat{\sigma}_z^1 \hat{\sigma}_z^2 \hat{\sigma}_z^2 = -1 \end{aligned} \quad (3.19)$$

Computing the product of the observables in the third row, i.e.:

$$\begin{aligned} \{\hat{\sigma}_x^1 \hat{\sigma}_y^2\} \{\hat{\sigma}_x^2 \hat{\sigma}_y^1\} \{\hat{\sigma}_z^1 \hat{\sigma}_z^2\} &= \hat{\sigma}_x^1 \underbrace{\hat{\sigma}_y^2 \hat{\sigma}_x^2}_{=-i\hat{\sigma}_z^2} \hat{\sigma}_y^1 \{\hat{\sigma}_z^1 \hat{\sigma}_z^2\} = \underbrace{\hat{\sigma}_x^1 \hat{\sigma}_y^1}_{=i\hat{\sigma}_z^1} \underbrace{\{-i\hat{\sigma}_z^2\}}_{\text{commute } so \leftrightarrow} \{\hat{\sigma}_z^1 \hat{\sigma}_z^2\} \\ &= \underbrace{i\hat{\sigma}_z^1 \hat{\sigma}_z^1}_{=i} \underbrace{\{-i\hat{\sigma}_z^2\}}_{=-i} \{\hat{\sigma}_z^2\} = +1, \end{aligned} \quad (3.20)$$

If the product rule is applied to the value assignments made in the rows, then:

$$\begin{aligned} V_{\vec{\psi}}(\hat{\sigma}_x^1) V_{\vec{\psi}}(\hat{\sigma}_x^2) V_{\vec{\psi}}(\hat{\sigma}_x^1 \hat{\sigma}_x^2) &= V_{\vec{\psi}}(\hat{\sigma}_y^2) V_{\vec{\psi}}(\hat{\sigma}_y^1) V_{\vec{\psi}}(\hat{\sigma}_y^1 \hat{\sigma}_y^2) = \\ V_{\vec{\psi}}(\hat{\sigma}_x^1 \hat{\sigma}_y^2) V_{\vec{\psi}}(\hat{\sigma}_x^2 \hat{\sigma}_y^1) V_{\vec{\psi}}(\hat{\sigma}_x^1 \hat{\sigma}_y^1) &= +1 \end{aligned} \quad (3.21)$$

while the column identities require:

$$\begin{aligned} V_{\vec{\psi}}(\hat{\sigma}_x^1) V_{\vec{\psi}}(\hat{\sigma}_y^2) V_{\vec{\psi}}(\hat{\sigma}_x^1 \hat{\sigma}_y^2) &= V_{\vec{\psi}}(\hat{\sigma}_x^2) V_{\vec{\psi}}(\hat{\sigma}_y^1) V_{\vec{\psi}}(\hat{\sigma}_x^2 \hat{\sigma}_y^1) = +1 \\ V_{\vec{\psi}}(\hat{\sigma}_x^1 \hat{\sigma}_y^2) V_{\vec{\psi}}(\hat{\sigma}_y^1 \hat{\sigma}_y^2) V_{\vec{\psi}}(\hat{\sigma}_z^1 \hat{\sigma}_z^2) &= -1 \end{aligned} \quad (3.22)$$

However, it is easy to show that the 9 numbers  $V_{\vec{\psi}}$  cannot satisfy all 6 constraints: multiplying all 9 observables together gives 2 different results, a  $+1$  when it is done

row by row and a  $-1$  when it is done column by column:

$$V_{\vec{\psi}}(\hat{\sigma}_x^1)V_{\vec{\psi}}(\hat{\sigma}_x^2)V_{\vec{\psi}}(\hat{\sigma}_x^1\hat{\sigma}_x^2)\dots V_{\vec{\psi}}(\hat{\sigma}_z^1\hat{\sigma}_z^2) = +1 \quad (3.23)$$

$$V_{\vec{\psi}}(\hat{\sigma}_x^1)V_{\vec{\psi}}(\hat{\sigma}_y^2)V_{\vec{\psi}}(\hat{\sigma}_x^1\hat{\sigma}_y^2)\dots V_{\vec{\psi}}(\hat{\sigma}_z^1\hat{\sigma}_z^2) = -1 \quad (3.24)$$

There obviously is no consistent solution to eqs. 3.24 and 3.23 since they contain the same set of numbers, simply ordered differently. Therefore the values assigned to the observables cannot obey the same identities that the observables themselves obey,  $V_{\vec{\psi}}(F\{\hat{A}\}) \neq F\{V_{\vec{\psi}}(\hat{A})\}$ , and an HVT would have to assign values to observables in a way that depended on the choice of which of 2 mutually commuting sets of observables that were also chosen to measure, i.e. the values assigned are contextual.

### 3.2 ABL, VAA and BKS nonets

Following Vaidman, Albert, and Aharonov (VAA) [4], Mermin showed how to assign a definite value to a single measurement of any one of the nine observables of a BKS nonet [27]. He then generalized this to a definite assignment to any one of 16 observables from the sets of nonets and showed that this assignment cannot be done if one attempts to measure (or ascertain) 2 or more of the observables belonging to the nonets. He left open the question as to the physical reason for this, stating “I find this intriguing” [27] To address this, we present a physical reason to demonstrate why the VAA scheme cannot be applied to 2 or more measurements by showing that the 2 measurements interfere with each other given the necessary PPSs. This can be seen to be a consequence of TSQM: some assignment of eigenvalues to operators are based on just one of the two vectors (i.e. either the pre- *or* the post-selected vector) while some assignment of eigenvalues are based on both vectors (i.e. both the pre-selected *and* the post-selected vectors - what we call diagonal-PPS). In this picture, it is the utilization of more than one PPS and the subsequent interference between them that explains the violation of the product rule and thus the physical source of the “contextuality” (provided that the contradictory statements are obtained by measuring the product of observables, rather than by measuring the observables individually, and this is always the case). When assignments are not made in the diagonal-PPS sense, then sets of commuting observables which are determined entirely by just one vector **satisfy** the BKS function condition  $V_{\vec{\psi}}(F\{\hat{A}\}) = F\{V_{\vec{\psi}}(\hat{A})\}$ . Sets of commuting observables which are assigned values in the diagonal-PPS sense by using information from both vectors **do not** satisfy the BKS function condition because they violate the product rule, and can disturb each other.

#### 3.2.1 Ascertaining the results of any one of the 9 observables:

We begin our analysis by considering specific examples of PPS configurations. We then utilize ABL [2] and show that choosing different post-selections change which triplet of observables violate the product rule. Consider first a pre-selection of  $\hat{\sigma}_x^1 = 1$  and  $\hat{\sigma}_x^2 = 1$  and a post-selection of  $\hat{\sigma}_y^1 = 1$  and  $\hat{\sigma}_y^2 = 1$  (see fig. 3.a). In this case,

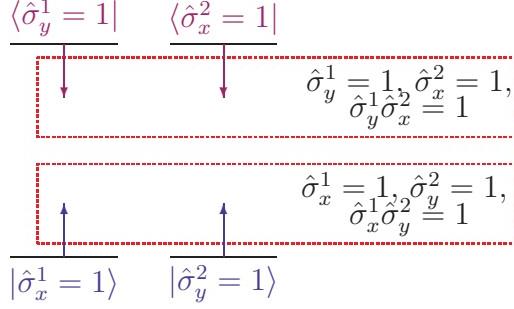
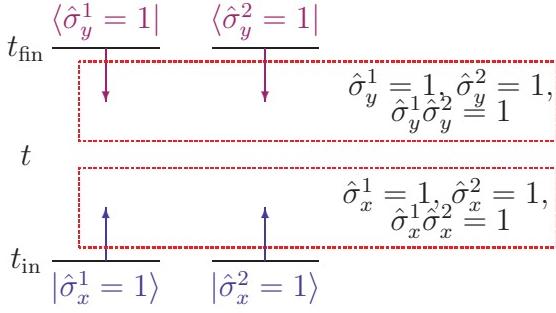


Figure 3: PPS states for Mermin example.

it is easy to see that we can ascertain with certainty any one of the following values  $\hat{\sigma}_x^1 = 1$ ,  $\hat{\sigma}_x^2 = 1$ ,  $\hat{\sigma}_y^1 = 1$  or  $\hat{\sigma}_y^2 = 1$ . We also know that we will obtain definite values of +1 if we measure any one of the following products of observables:  $\hat{\sigma}_x^1\hat{\sigma}_x^2 = 1$  and  $\hat{\sigma}_y^1\hat{\sigma}_y^2 = 1$ , so we must also obtain  $\hat{\sigma}_x^1\hat{\sigma}_x^2\hat{\sigma}_y^1\hat{\sigma}_y^2 = +1$ .

$$\langle \hat{\sigma}_x^1\hat{\sigma}_x^2\hat{\sigma}_y^1\hat{\sigma}_y^2 \rangle = \langle \hat{\sigma}_y^1 = 1 | \langle \hat{\sigma}_y^2 = 1 | \hat{\sigma}_x^1\hat{\sigma}_x^2\hat{\sigma}_y^1\hat{\sigma}_y^2 | \hat{\sigma}_x^1 = 1 \rangle | \hat{\sigma}_x^2 = 1 \rangle = +1 \quad (3.25)$$

In addition, we obtain the same results if we switch the sequence of  $\hat{\sigma}_x^1\hat{\sigma}_x^2$  and  $\hat{\sigma}_y^1\hat{\sigma}_y^2$  because they commute. Any *one* of the other observables in fig. 2 (i.e.  $\hat{\sigma}_x^1\hat{\sigma}_y^2$  and  $\hat{\sigma}_x^2\hat{\sigma}_y^1$ ) can also be ascertained with certainty given this PPS. Finally, from the product of the 3 observables in column 3 ( $\{\hat{\sigma}_x^1\hat{\sigma}_x^2\}\{\hat{\sigma}_y^1\hat{\sigma}_y^2\}\{\hat{\sigma}_z^1\hat{\sigma}_z^2\} = -1$  and  $\{\hat{\sigma}_x^1\hat{\sigma}_x^2\}\{\hat{\sigma}_y^1\hat{\sigma}_y^2\} = +1$ ), we can deduce that  $\hat{\sigma}_z^1\hat{\sigma}_z^2 = -1$ . Similar statements can be made for other PPSs. E.g. consider the pre-selection:  $\hat{\sigma}_x^1 = 1$  and  $\hat{\sigma}_y^2 = 1$  and a post-selection of  $\hat{\sigma}_y^1 = 1$  and  $\hat{\sigma}_x^2 = 1$  (see fig. 3.b) In this case, the measurement  $\hat{\sigma}_x^1\hat{\sigma}_y^2\hat{\sigma}_x^2\hat{\sigma}_y^1 = +1$  and we can deduce from the third row that  $\{\hat{\sigma}_z^1\hat{\sigma}_z^2\} = +1$ .

Thus, given just one PPS, any single observable can be assigned a definite value, even though  $\hat{\sigma}_z^1\hat{\sigma}_z^2$  is assigned different values in different PPS. It is precisely because of this connection between particular PPSs and different values for  $\hat{\sigma}_z^1\hat{\sigma}_z^2$  that the issue of contextuality arises when we consider products of these observables.

### 3.2.2 Ascertaining the results of products of the 9 observables:

In this section, we ask how many of the *products* of the 9 observables in fig. 2 can be ascertained together with certainty. For example, as stated in the previous section, the outcome for the product of the first two observables in column 3 of fig. 2 with the PPS of fig. 3.a is  $\sigma_x^1\sigma_x^2\sigma_y^1\sigma_y^2 = +1$ . However, if we measure the operators corresponding to the first 2 observables of row 3 in fig. 2, i.e.  $\hat{\sigma}_x^1\hat{\sigma}_y^2\hat{\sigma}_x^2\hat{\sigma}_y^1$ , given this particular PPS shown in fig. 3.a, then the sequence of measurements interfere with each other (as represented by the slanted ovals in figure 5.a). To see this, consider that  $\hat{\sigma}_x^1\hat{\sigma}_y^2\hat{\sigma}_x^2\hat{\sigma}_y^1$  corresponds to the sequence of measurements represented in figure 4.a. While the pre-selection of particle 2 is  $\hat{\sigma}_x^2 = 1$  at  $t_{in}$ , the next measurement after the pre-selection at  $t_2$  is for  $\hat{\sigma}_y^2$  and only *after* that a measurement of  $\hat{\sigma}_x^2$  is performed at  $t_3$ . Thus, there is no guarantee that the  $\hat{\sigma}_x^2$  measurement at  $t_3$  will give the same value as the pre-selected state of  $\hat{\sigma}_x^2 = 1$  or that the  $\hat{\sigma}_y^2$  measurement will give the same value as the post-selected state of  $\hat{\sigma}_y^2 = 1$ . In TSQM, this is due to the disturbance of the 2-vector boundary conditions which is created by the IM: the initial pre-selected vector  $\hat{\sigma}_x^2 = 1$  from  $t_{in}$  is “destroyed” when the  $\hat{\sigma}_y^2$  measurement at

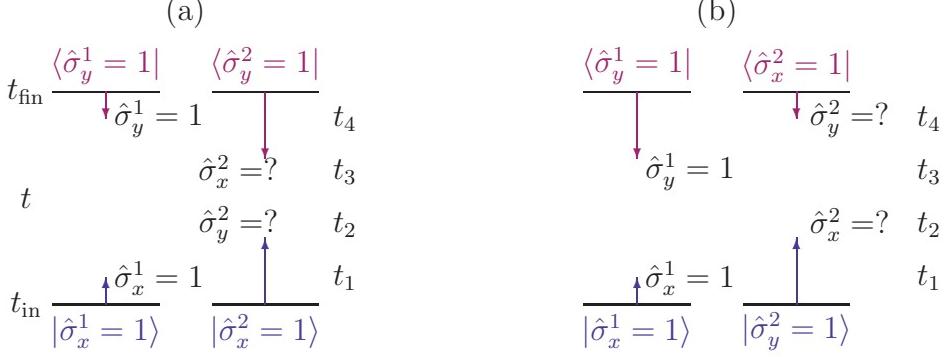


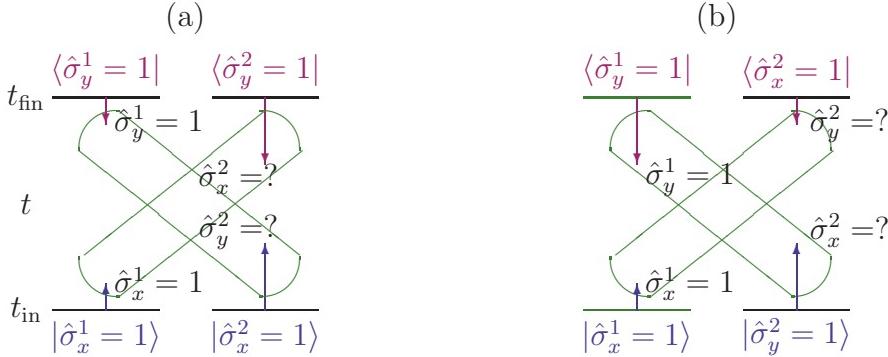
Figure 4: Time sequence of PPS measurements for Mermin example.

time  $t_2$  is performed and therefore cannot inform the later  $\hat{\sigma}_x^2$  measurement at time  $t_3$ . In other words, with the particular PPS given in fig. 3.a and 4.a, the operator,  $\hat{\sigma}_x^1 \hat{\sigma}_y^2 \hat{\sigma}_x^2 \hat{\sigma}_y^1$  depends on information from both the pre-selected vector  $\hat{\sigma}_x^1 = 1$ ,  $\hat{\sigma}_x^2 = 1$  and the post-selected vector  $\hat{\sigma}_y^1 = 1$ ,  $\hat{\sigma}_y^2 = 1$  in a diagonal-PPS sense. We call this diagonal-PPS because a line connecting  $\hat{\sigma}_x^1(t_1)$  with  $\hat{\sigma}_x^2(t_3)$  will be diagonal or will cross the line connecting  $\hat{\sigma}_y^2(t_2)$  with  $\hat{\sigma}_y^1(t_4)$ , where  $t_{\text{in}} < t_1 < t_2 \dots < t_{\text{fin}}$ , see fig. 5.a).

These results can also be seen in an actual measurement situation, we consider interaction Hamiltonians with coupling terms  $\hat{\sigma}_x^1 \delta(t - t_1)$ ,  $\hat{\sigma}_x^2 \delta(t - t_2)$ ,  $\hat{\sigma}_y^2 \delta(t - t_2)$  and  $\hat{\sigma}_y^1 \delta(t - t_1)$ :

$$\begin{aligned} & \langle \hat{\sigma}_y^1 = 1 | \langle \hat{\sigma}_y^2 = 1 | e^{i\hat{Q}_1 \hat{\sigma}_x^1 \hat{\sigma}_x^2} e^{i\hat{Q}_2 \hat{\sigma}_y^1 \hat{\sigma}_y^2} | \hat{\sigma}_x^1 = 1 \rangle | \hat{\sigma}_x^2 = 1 \rangle \\ &= \langle \hat{\sigma}_y^1 = 1 | \langle \hat{\sigma}_y^2 = 1 | e^{i\hat{Q}_2 \hat{\sigma}_y^1 \hat{\sigma}_y^2} e^{i\hat{Q}_1 \hat{\sigma}_x^1 \hat{\sigma}_x^2} | \hat{\sigma}_x^1 = 1 \rangle | \hat{\sigma}_x^2 = 1 \rangle = 1 \end{aligned} \quad (3.26)$$

Since  $e^{i\hat{Q}_1 \hat{\sigma}_x^1 \hat{\sigma}_x^2}$  commutes with  $e^{i\hat{Q}_2 \hat{\sigma}_y^1 \hat{\sigma}_y^2}$ , they can be interchanged and thus the same outcome of +1 is obtained. However, for the other observables  $\langle \hat{\sigma}_y^1 = 1 | \langle \hat{\sigma}_y^2 = 1 | e^{i\hat{Q}_1 \hat{\sigma}_x^1 \hat{\sigma}_x^2} e^{i\hat{Q}_2 \hat{\sigma}_y^1 \hat{\sigma}_y^2} | \hat{\sigma}_x^1 = 1 \rangle | \hat{\sigma}_x^2 = 1 \rangle$  the opposite eigenvalue is obtained (even though they commute) i.e.  $\hat{\sigma}_x^1 \hat{\sigma}_y^2 \hat{\sigma}_x^2 \hat{\sigma}_y^1$  will give an outcome of -1 given the PPS of fig. 4.a even though separately  $\hat{\sigma}_x^1 \hat{\sigma}_y^2 = +1$  and  $\hat{\sigma}_x^2 \hat{\sigma}_y^1 = +1$ . This is thus a violation of the product rule (see fig. 5.a). This diagonal-PPS phenomenon can be generalized to functions that are polynomials of products of observables with the proper ordering (i.e. no mixing or sandwiching).

Figure 5: a) Measurement of  $\hat{\sigma}_x^1 \hat{\sigma}_y^2 \hat{\sigma}_x^2 \hat{\sigma}_y^1$  is diagonal, b) measurement of  $\hat{\sigma}_x^1 \hat{\sigma}_x^2 \hat{\sigma}_y^1 \hat{\sigma}_y^2$  is diagonal

To summarize this sub-section, given the PPS of fig. 5.a, the subset of observables circled in fig. 6.a (and the products of those circled observables) can be assigned eigenvalues in a way that satisfies the function relation requirement eq. 1.1. But, the product of the other observables (e.g.  $\hat{\sigma}_x^1 \hat{\sigma}_y^2$  and  $\hat{\sigma}_x^2 \hat{\sigma}_y^1$ ) can only be ascertained (given this particular PPS) using information from *both* the pre- and post-selected vector in a diagonal sense (see fig. 5.a), and will thus violate the product rule. With the PPS of fig. 5.b, the subset in fig. 6.b (and the relevant products of observables) can be assigned eigenvalues in a way that satisfies the function relation requirement eq. 1.1. But, the product of the other observables  $\hat{\sigma}_x^1 \hat{\sigma}_x^2$  and  $\hat{\sigma}_y^1 \hat{\sigma}_y^2$ , i.e.  $\hat{\sigma}_x^1 \hat{\sigma}_x^2 \hat{\sigma}_y^1 \hat{\sigma}_y^2$  violate the function rule eq. 1.1.

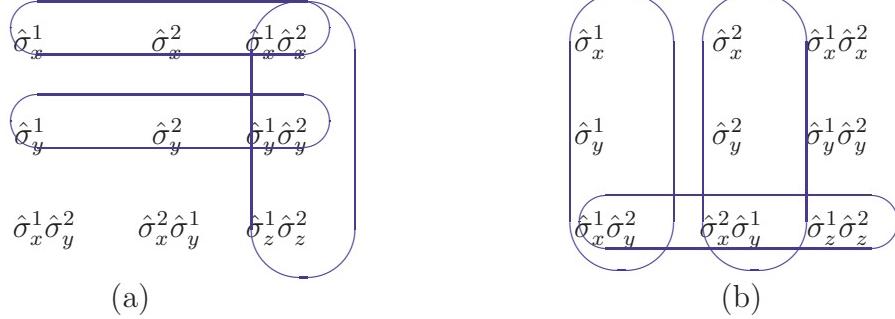


Figure 6: Products of observables that are not disturbed, a) given the PPS of fig. 2.a, and b) given the PPS of fig. 2.b

### 3.2.3 Ascertaining the results of any one of 16 observables through the generalized state:

As was explained in §3.2.1, definite results for any *one* of the intermediate measurements can be obtained for several different PPSs which were complete measurements (and thus describable by a wavefunction). In addition, some triplets (products of observables) can also be ascertained (see fig. 6) given a particular pre- or post-selection. However, for the most general setup considered in this section, no 2 products of observables can be ascertained. The general setup considers superpositions of these PPSs. In fact, this is required in order to ascertain any one of the 16 observables in Mermin's successful generalization of VAA (it is also required to ascertain any one of the 9 for general PPSs). Following VAA [4], Mermin showed that the way that any one of the 16 values of the 7 BKS nonets (e.g. fig. 2 is one of them, the others given by  $\sigma_\nu \sigma_\mu$ )

can be ascertained with certainty is by entangling the 2-particle system representing the 4D BKS nonet (represented by the pre and post-selection of fig. 3.a and fig. 3.b, etc., and labeled as  $|i\rangle$ ) with another system, i.e. an ancilla (represented here by  $|\Psi_i\rangle$  and  $|\Phi_i\rangle$ ). The nonet is prepared at  $t_{in}$  by correlating it with a set of states of an ancilla:

$$|A\rangle = \sum_i \frac{1}{\sqrt{N}} |\Phi_i\rangle |i\rangle \quad (3.27)$$

Then the ancilla is “guarded” so there are no interactions with the ancilla during the

time  $(t_{\text{in}}, t_{\text{fin}})$ . At  $t_{\text{fin}}$  we post select on the particle and ancilla and obtain the state:

$$|\Omega\rangle = \frac{1}{\sqrt{N}} \sum_i |\Psi_i\rangle |i\rangle \quad (3.28)$$

If we are successful in obtaining this state for the post-selection, then the state of the system is described in the intermediate time by the entangled state (see figure 7) [6, 9]

$$\Psi = \sum_i \alpha_i \langle \Psi_i || \Phi_i \rangle \quad (3.29)$$

Figure 7: Generalized State: superpositions of 2-vectors given by eq. 3.29.

For general PPSs, we use multiple sets of boundary conditions given by fig. 3.a and fig. 3.b, etc., to get an entangled state represented by fig. 8 (where for simplicity we have taken the states of the ancilla to be an orthonormal set,  $\mathbf{f}^\mu$ ). Mermin then presented an elegant method to determine the states of the ancilla necessary to produce the effect: he selected a definite representation for the 2-particle spins, performed a projection,  $\hat{\mathbf{P}}$ , onto the subspace given by the nonet, and solved under rotation for the state of the ancilla. That is,  $\langle \Omega | \hat{\mathbf{P}} | A \rangle = 0$  for “all but a single one of the projections associated with...” [27] the observable of the nonet that is to be ascertained with certainty.

$$\begin{aligned} t_{\text{fin}} & \left\{ \begin{array}{c} \hat{\sigma}_y^1 = 1 \\ \hat{\sigma}_y^2 = 1 \end{array} \right\} \otimes |\mathbf{f}^0\rangle + \left\{ \begin{array}{c} \hat{\sigma}_y^1 = 1 \\ \hat{\sigma}_x^2 = 1 \end{array} \right\} \otimes |\mathbf{f}^1\rangle + \dots \\ t & \\ t_{\text{in}} & \left\{ \begin{array}{c} \hat{\sigma}_x^1 = 1 \\ \hat{\sigma}_x^2 = 1 \end{array} \right\} \otimes |\mathbf{f}^0\rangle + \left\{ \begin{array}{c} \hat{\sigma}_x^1 = 1 \\ \hat{\sigma}_y^2 = 1 \end{array} \right\} \otimes |\mathbf{f}^1\rangle + \dots \end{aligned}$$

Figure 8: Generalized state for BKS nonets

As Mermin proved, all 4 components of this generalized state are necessary (i.e. we only determine a Bell-state at  $t_{\text{fin}}$  on the ancilla rather than make a projection onto any given component  $\mathbf{f}^\mu$ ) to ascertain a definite answer to any one of the individual observables.

### 3.2.4 A physical reason for restrictions on these assignments

We have suggested a physical reason based on TSQM and PPS for the 2 different values for  $\hat{\sigma}_z^1 \hat{\sigma}_z^2$ . This points to a physical reason why no 2 measurements can be

ascertained with certainty in the intermediate<sup>3</sup> time: all sets of boundary conditions are needed (those corresponding to both 3.a and 3.b, etc.) in order to ascertain with certainty the value of any one of the 16 observables, as represented by fig. 8. However, when the first observable is ascertained, then it will depend on both the pre- and the post-selection measurement (i.e. it will be diagonal-PPS) in 2 of the 4 components of the generalized state (see §3.2.2) and will collapse the entire configuration onto a subset of the PPSs, thereby disturbing the terms of the generalized state. Given any pair of measurements, there will always be a diagonal-PPS situation when all 4 components of the generalized state are considered. This can be seen by comparing figs. 6.a and 6.b and noting that any 2 observables will not be circled in both. Therefore, since we cannot be sure that the entire setup (see fig. 8) is not disturbed, we cannot ascertain with certainty the outcome for any one of the 16 observables for the second measurement. Furthermore, this arrangement is the maximal correlation that can be performed (i.e. the 4D state of 2 spins can be maximally correlated to another 4D system as performed here), and thus we cannot create an even more sophisticated situation with additional ancillas. We have thus given a physical picture for Mermin’s “intriguing” question: there will always be a diagonal situation for any 2 observables.

### 3.3 Non-classical Weak Values for the 4D BKS nonets

We can now clarify Mermin’s statement: “Alice’s other two ‘results’ have nothing to do with any properties of the particle or the results of any measurement actually performed.” [27] While it is certainly true that these “other results” cannot be ascertained simultaneously in terms of an IM (as was demonstrated in §3.2.4 and by Mermin), they **can** be measured simultaneously through WMs.

The route to an easy calculation of WVs can be established from Mermin’s description of VAA’s accomplishment: “Alice’s list gives the observed result for the measurement Bob actually made and had he measured anything else it would have given the result he observed.” [27] This provides a direct route to WMs through theorem 2: WMs will produce the identical result as predicted for the IM since the IM results are definite. Thus, the other ‘results’ are related to properties of the particle and can be simultaneously measured.

We can also obtain non-classical results in this example (similar to the 3-box-paradox), by first re-writing the observables of fig. 2 in terms of 3 spin components for 2 “virtual” particles: for the first particle,  $\hat{\mathbf{S}}_3^1 \equiv \hat{\sigma}_x^1 \hat{\sigma}_y^2$ ,  $\hat{\mathbf{S}}_2^1 \equiv \hat{\sigma}_x^1 \hat{\sigma}_z^2$ , and  $\hat{\mathbf{S}}_1^1 \equiv \hat{\sigma}_x^2$ ; for particle 2, components which commute with particle 1 are  $\hat{\mathbf{S}}_3^2 \equiv \hat{\sigma}_x^2 \hat{\sigma}_y^1$ ,  $\hat{\mathbf{S}}_2^2 \equiv \hat{\sigma}_z^1 \hat{\sigma}_x^2$ , and  $\hat{\mathbf{S}}_1^2 \equiv \hat{\sigma}_x^1$ . We can observe a non-classical WV by noting that  $\hat{\sigma}_x^1 \hat{\sigma}_y^2 \hat{\sigma}_x^2 \hat{\sigma}_y^1 = -1$  given the PPS of fig. 4.a even though separately  $\hat{\sigma}_x^1 \hat{\sigma}_y^2 = +1$  and  $\hat{\sigma}_x^2 \hat{\sigma}_y^1 = +1$ , i.e. a violation of the product rule and thus a diagonal situation (see fig. 5.a). Thus WMs must yield the same outcomes, i.e.  $\hat{N}(\hat{\mathbf{S}}_3^1)_w = +1$  and  $\hat{N}(\hat{\mathbf{S}}_3^2)_w = +1$  but

---

<sup>3</sup>If we were considering a single PPS, as discussed in §3.2.1 and §3.2.2, then *some* pairs of products of observables can be ascertained with certainty, but not any 2 pairs. No 2 products can be ascertained in the case of 16 observables and some 2 pairs of products cannot be ascertained for the 9 observables.

$\hat{N}(\hat{S}_3^1 \hat{S}_3^2)_w = -1$ , a non-classical result. To analyze these results, we define the following pair occupation operators:

- $\hat{N}_{++}$  the projector on the state  $\hat{S}_3^1 = 1$ , and  $\hat{S}_3^2 = 1$
- $\hat{N}_{+-}$  the projector on the state  $\hat{S}_3^1 = 1$ , and  $\hat{S}_3^2 = -1$
- $\hat{N}_{-+}$  the projector on the state  $\hat{S}_3^1 = -1$ , and  $\hat{S}_3^2 = 1$
- $\hat{N}_{--}$  the projector on the state  $\hat{S}_3^1 = -1$ , and  $\hat{S}_3^2 = -1$ .

We can relate these measurements to the 3 box example of §2, but in this case we have 2 boxes and 2 particles:  $\hat{N}_{++}$  means the number of times that particle 1 and particle 2 are in the first box,  $\hat{N}_{+-}$  means the number of times that particle 1 is in the first box and particle 2 is in the second box, etc.

The WV of the projection operator  $(1 - \hat{S}_3^1)(1 - \hat{S}_3^2)$  is  $-\frac{1}{2}$ , a non-classical result. From Theorems 1 and 2, we can deduce the following: the different ways to obtain  $\hat{N}(\hat{S}_3^2) = +1$  are given by  $\hat{N}_{++}$  (i.e.  $\hat{S}_3^1 = 1$ , and  $\hat{S}_3^2 = 1$ ) and  $\hat{N}_{-+}$  (i.e.  $\hat{S}_3^1 = -1$ , and  $\hat{S}_3^2 = 1$ ) and therefore we can deduce that:

$$\hat{N}_{*+} \equiv \hat{N}_{++} + \hat{N}_{-+} = 1 \quad (3.30)$$

In terms of the box analogy, this is how many ways that particle 2 can be found in box 1. Also  $\hat{N}(\hat{S}_3^1) = +1$ , and thus  $\hat{N}(\hat{S}_3^1) \neq -1$  (i.e. how many ways can particle 1 be found in box 1). This is characterized by:

$$\hat{N}_{+*} \equiv \hat{N}_{++} + \hat{N}_{+-} = 1 \quad (3.31)$$

From  $\hat{S}_3^1 \hat{S}_3^2 = -1$  (which again means that both particles cannot be found in the same box), it cannot be that  $\hat{S}_3^2 \hat{S}_3^1 = +1$  and thus  $\hat{S}_3^2$  and  $\hat{S}_3^1$  must be opposite in sign. This can be characterized by:

$$\hat{N}_{++} + \hat{N}_{--} = 0 \quad (3.32)$$

Furthermore, since eq. 3.30 equals eq. 3.31, we can deduce:

$$\hat{N}_{+-} = \hat{N}_{-+} \quad (3.33)$$

Subtracting eq. 3.32 from the following identity

$$\hat{N}_{++} + \hat{N}_{--} + \hat{N}_{+-} + \hat{N}_{-+} = +1. \quad (3.34)$$

we obtain:

$$\hat{N}_{+-} + \hat{N}_{-+} = 1 \quad (3.35)$$

Eq. 3.35 implies that the 2 particles are never in the same box. From eq. 3.33 and eq. 3.35, we can deduce:

$$\hat{N}_{-+} = \hat{N}_{+-} = \frac{1}{2}. \quad (3.36)$$

Substituting this value into eq. 3.31, we can deduce that:

$$\hat{N}_{++} = \frac{1}{2} \quad (3.37)$$

Finally, substituting this into eq. 3.32, we can deduce:

$$\hat{\mathbf{N}}_{--} = -\frac{1}{2}. \quad (3.38)$$

As shown in [11], all these statements can be measured simultaneously through WMs and will yield:

$$(\hat{\mathbf{N}}_{-+})_w = (\hat{\mathbf{N}}_{+-})_w = (\hat{\mathbf{N}}_{++})_w = \frac{1}{2}. \quad (3.39)$$

while:

$$(\hat{\mathbf{N}}_{--})_w = -\frac{1}{2} \quad (3.40)$$

In other words, if a WM is performed on the number of times that particle 1 is in the first box and particle 2 is in the first box, then the result is the non-classical result  $-\frac{1}{2}$ . Thus, the way that the 2 seemingly contradictory statements  $\hat{\sigma}_z^1 \hat{\sigma}_z^2 = \pm 1$  can weakly “peacefully co-exist” (to paraphrase Abner Shimony) is that a WV goes outside the spectrum of possible eigenvalues, i.e. eq. 3.40.

In summary, we see again that WMs give an empirical manifestation of BKS:

- the BKS “contradiction” here is that  $\hat{\sigma}_x^1 \hat{\sigma}_y^2 \hat{\sigma}_x^2 \hat{\sigma}_y^1 = -1$  (given the PPS of fig. 4.a) even though separately  $\hat{\sigma}_x^1 \hat{\sigma}_y^2 = +1$  and  $\hat{\sigma}_x^2 \hat{\sigma}_y^1 = +1$
- these 3 outcomes can be measured weakly without contradiction because the product of WVs is not equal to the WV of the product
- if BKS were not correct and a noncontextual-HVT were possible, then the product rule should be satisfied and an IM of  $\hat{\sigma}_x^1 \hat{\sigma}_y^2 \hat{\sigma}_x^2 \hat{\sigma}_y^1$  should yield  $+1$ . This leads to an immediate contradiction because:
  - by theorem 2, the WV must be equal to the ideal result
  - but, this would be inconsistent with an actual WM which will register  $(\hat{\mathbf{N}}_{--})_w = -\frac{1}{2}$ ,
- therefore, BKS is empirically consistent with WMs

We have thus given a physical explanation for why an IM cannot reveal these values, while WMs can reveal these values. Thus, with WMs, the BKS “contradiction” still exists (i.e. a noncontextual-HVT cannot reproduce QM), yet now it can also be measured. In other words, we have physically shown how to obtain

- $+1$  for the product of all nine observables when this is performed in the sequence of the rows of fig. 2
- $-1$  for the product of all nine observables when this is performed in the sequence of the columns of fig. 2

(assuming that the system is PPS and measured weakly). The ambiguity in determining whether  $\hat{\sigma}_z^1 \hat{\sigma}_z^2 = +1$  is obtained or  $\hat{\sigma}_z^1 \hat{\sigma}_z^2 = -1$  is obtained gets shifted to the

ambiguity of determining which set of boundary conditions is obtained, i.e. it is now a physical property of the system.

A modification to the experimental setup suggested by [39] could be used to test the predictions made in this paper. This setup considers a 4-D Hilbert space represented by two 2-D subsystems, e.g. the path and polarization of a single photon. [39] find a BKS contradiction for a particular entangled state. If the transverse positions of each photon is used as the pointer, then WMs can be obtained in [39] with small transverse displacements. The postselected photon distribution then determines the size of these displacements. Using the technique of [10], WMs can be performed if an optical glass is slightly tilted so that the photon's position is uncertain to within the width of the beam. As shown in the next section, the WV<sub>s</sub> calculated in §3.3 (e.g. eqs. 3.40 and 3.39) are identical to WV<sub>s</sub> for EPR entanglement [30] and thus entanglement in a pre-selected state of 2 particles is isomorphic to an entanglement in our 2 virtual particles.

### 3.4 Non-classical weak values for EPR and Peres/BKS

We shall now show that WV<sub>s</sub> can show new connections between BKS and EPR. Eq. 3.40 and 3.39 give the same result as calculating WV<sub>s</sub> for EPR entanglement [30] and thus there is an interesting new kind of isomorphism between the problems of WV<sub>s</sub> in BKS nonets and EPR. We can also consider interesting manifestations of the “contextuality” in these situations by making separate measurements on the ancilla. Consider a pre-selected state which is entangled between one of the particles of the nonet and the ancilla (where, following Mermin, we have chosen the ancilla to be an orthonormal set  $|\mathbf{f}^1\rangle$ ):

$$|A\rangle = \frac{1}{\sqrt{2}}(|\hat{\sigma}_z^1 = +1\rangle|\mathbf{f}^0 = +1\rangle - |\hat{\sigma}_z^1 = -1\rangle|\mathbf{f}^0 = -1\rangle) \quad (3.41)$$

If we consider product state post-selections, e.g. with  $\hat{\sigma}_x^1 = 1$  and  $\mathbf{f}^0 = +1$ , then we know that an ideal measurement of  $\hat{\sigma}_x^1\mathbf{f}^0$  must yield

$$\hat{\sigma}_x^1\mathbf{f}^0 = 1 \quad (3.42)$$

The pre-selected state also yields<sup>4</sup>

$$\hat{\sigma}_x^1\mathbf{f}^0 + \mathbf{f}^1\hat{\sigma}_z^1 = 0. \quad (3.43)$$

Thus

$$\begin{aligned} & \{\hat{\sigma}_x^1\mathbf{f}^0 + \mathbf{f}^1\hat{\sigma}_z^1\}\{| \uparrow_z^1\rangle|\mathbf{f}^0 = -1\rangle - | \downarrow_z^1\rangle|\mathbf{f}^0 = +1\rangle\} = \\ & \{-| \downarrow_z^1\rangle|\mathbf{f}^0 = -1\rangle + | \uparrow_z^1\rangle|\}\} - \{| \uparrow_z^1\rangle|\} - | \downarrow_z^1\rangle|\mathbf{f}^0 = -1\rangle\} = 0 \end{aligned} \quad (3.44)$$

---

<sup>4</sup>This is easy to see because  $\mathbf{f}^1\hat{\sigma}_z^1| \uparrow_z^1\rangle|\mathbf{f}^0 = -1\rangle \rightarrow | \uparrow_z^1\rangle|\mathbf{f}^0 = +1\rangle$  and  $\hat{\sigma}_x^1\mathbf{f}^0| \uparrow_z^1\rangle|\mathbf{f}^0 = -1\rangle \rightarrow -| \downarrow_z^1\rangle|\mathbf{f}^0 = -1\rangle$  and  $\mathbf{f}^1\hat{\sigma}_z^1| \downarrow_z^1\rangle|\mathbf{f}^0 = +1\rangle \rightarrow -| \downarrow_z^1\rangle|\mathbf{f}^0 = -1\rangle$  and  $\hat{\sigma}_x^1\mathbf{f}^0| \downarrow_z^1\rangle|\mathbf{f}^0 = +1\rangle \rightarrow | \uparrow_z^1\rangle|\mathbf{f}^0 = +1\rangle$ .

We can deduce outcomes for un-performed measurements of  $\hat{\sigma}_z^1$ :

$$\hat{\sigma}_z^1 = -1. \quad (3.45)$$

(we have assumed that the ancilla is not disturbed to obtain this conclusion). We can also deduce outcomes for un-performed measurements of  $\mathbf{f}^1$  (assuming that the first particle is not disturbed):

$$\mathbf{f}^1 = -1. \quad (3.46)$$

Once again, we see a violation of the product rule [29]: from eq. 3.44 we deduce that  $\mathbf{f}^1\hat{\sigma}_z^1 = -1$ , but if constructed individually from eqs. 3.46 and 3.45,  $\mathbf{f}^1\hat{\sigma}_z^1 = +1$ . This “conclusion,” however, relies on counterfactual statements, since not all the required measurements eqs. 3.43-3.46 can be performed simultaneously without disturbing each other. However, we **can** perform WMs on all these statements simultaneously (see §3.3).

WVs in the EPR situation can also be seen in the instant example if we consider this PPS (instead of a Bell-state, we measure a definite state of the ancilla). We see the identical non-classical results as seen in the previous section. First, we define the following projectors

- $N_{++}$  the projector on the state  $\hat{\sigma}_z^1 = 1$ , and  $\mathbf{f}^1 = 1$
- $N_{+-}$  the projector on the state  $\hat{\sigma}_z^1 = 1$ , and  $\mathbf{f}^1 = -1$
- $N_{-+}$  the projector on the state  $\hat{\sigma}_z^1 = -1$ , and  $\mathbf{f}^1 = 1$
- $N_{--}$  the projector on the state  $\hat{\sigma}_z^1 = -1$ , and  $\mathbf{f}^1 = -1$ .

From Theorem 1 and 2, we can deduce the following: from the post-selection  $\hat{\sigma}_x^1 = +1$ , we can deduce that  $\mathbf{f}^1 = -1$ , i.e. eq. 3.46. The different ways to obtain  $\mathbf{f}^1 = -1$  are given by  $N_{+-}$  (i.e.  $\hat{\sigma}_z^1 = 1$ , and  $\mathbf{f}^1 = -1$ ) and  $N_{--}$  (i.e.  $\hat{\sigma}_z^1 = -1$ , and  $\mathbf{f}^1 = -1$ ) and therefore we can deduce the conservation relationship:

$$N_{+-} + N_{--} = 1 \quad (3.47)$$

Now  $\hat{\sigma}_z^1 = -1$ , and thus  $\hat{\sigma}_z^1 \neq +1$  is characterized by:

$$N_{+*} \equiv N_{++} + N_{-+} = 0 \quad (3.48)$$

In terms of the box analogy, this is how many ways that particle 1 can be found in box 1. In addition, if  $\mathbf{f}^1 = -1$  then  $\mathbf{f}^1 \neq +1$  thereby giving:

$$N_{++} + N_{-+} = 0 \quad (3.49)$$

That is, in how many ways can particle 2 be found in box 2. Furthermore, since eq. 3.49 equals eq. 3.48, we can deduce:

$$N_{+-} = N_{-+} \quad (3.50)$$

From the post-selection, we know that an ideal measurement of  $\hat{\sigma}_x^1\mathbf{f}^0$  will yield  $+1$ , therefore, using eq. 3.43, we can deduce that  $\mathbf{f}^1\hat{\sigma}_z^1 = -1$  (which means that both

particles cannot be found in the same box), i.e.

$$\underbrace{\hat{\sigma}_x^1 \mathbf{f}^0}_{=+1} + \underbrace{\mathbf{f}^1 \hat{\sigma}_z^1}_{\Rightarrow = -1} = 0. \quad (3.51)$$

Next we ask how to obtain  $\mathbf{f}^1 \hat{\sigma}_z^1 = -1$ . To obtain this, it cannot be that  $\mathbf{f}^1 \hat{\sigma}_z^1 = +1$  and thus  $\mathbf{f}^1$  and  $\hat{\sigma}_z^1$  must be opposite in sign. This can be characterized by:

$$N_{++} + N_{--} = 0 \quad (3.52)$$

Subtracting eq. 3.52 from the following identity

$$N_{++} + N_{--} + N_{+-} + N_{-+} = +1. \quad (3.53)$$

we obtain:

$$N_{+-} + N_{-+} = 1 \quad (3.54)$$

From eq. 3.50 and eq. 3.54, we can deduce:

$$N_{-+} = N_{+-} = \frac{1}{2}. \quad (3.55)$$

Plugging this value into eq. 3.48, we can deduce that:

$$N_{++} = -\frac{1}{2} \quad (3.56)$$

Finally, plugging this into eq. 3.52, we can deduce:

$$N_{--} = \frac{1}{2}. \quad (3.57)$$

Entanglement in a pre-selected state of 2 particles is isomorphic to an entanglement in our 2 virtual particles. Thus, if we look at the right variables, then the BKS setup can be seen to be related to EPR.

The situation analyzed above is general and also points to the Peres/BKS-example [49]: in summary, consider a pre-selected state  $|\Psi_{EPR}\rangle = \frac{1}{\sqrt{2}}(|\uparrow_z^1\rangle|\downarrow_z^2\rangle - |\downarrow_z^1\rangle|\uparrow_z^2\rangle)$  for which the following identity holds  $\hat{\sigma}_x^1 \hat{\sigma}_y^2 + \hat{\sigma}_x^2 \hat{\sigma}_y^1 = 0$ . In addition,  $|\Psi_{EPR}\rangle$  is also an eigenvector (with eigenvalue  $-1$ ) of the following operators  $\hat{\sigma}_x^1 \hat{\sigma}_x^2, \hat{\sigma}_y^1 \hat{\sigma}_y^2, \hat{\sigma}_z^1 \hat{\sigma}_z^2$ . It's easy to see that a noncontextual-HVT cannot assign values consistent with these operator relations:

$$V_{\vec{\psi}}(\hat{\sigma}_x^1 \hat{\sigma}_x^2) = V_{\vec{\psi}}(\hat{\sigma}_y^1 \hat{\sigma}_y^2) = V_{\vec{\psi}}(\hat{\sigma}_z^1 \hat{\sigma}_z^2) = -1 \quad (3.58)$$

Consider the commuting observables  $\hat{A}_1 = \hat{\sigma}_x^1 \hat{\sigma}_y^2$  and  $\hat{A}_2 = \hat{\sigma}_x^2 \hat{\sigma}_y^1$ . We know that

$$\hat{A}_1 \hat{A}_2 = \hat{\sigma}_x^1 \hat{\sigma}_y^2 \hat{\sigma}_x^2 \hat{\sigma}_y^1 = \hat{\sigma}_z^2 \hat{\sigma}_z^1 = -1 \quad (3.59)$$

in the singlet state. The assumption of non-contextuality is that value assignments can be made to eq. 3.59 even when these assignments are taken from a different

context: e.g. assigning values from eq. 3.58, we obtain:

$$V_{\vec{\psi}}(\hat{\sigma}_x^1)V_{\vec{\psi}}(\hat{\sigma}_y^2)V_{\vec{\psi}}(\hat{\sigma}_x^2)V_{\vec{\psi}}(\hat{\sigma}_y^1) = (-1)(-1) = +1 \quad (3.60)$$

but experimentally, we obtain  $-1$  from eq. 3.59, a contradiction. Thus a noncontextual HVT is impossible. However, now we can probe this state by post-selections and obtain  $\hat{\sigma}_x^1\hat{\sigma}_y^2 = -1$  even though  $\hat{\sigma}_x^1 = -1$  and  $\hat{\sigma}_y^2 = -1$ . This, again, can only be done with PPSs since in a pre-selected only system, for 2 commuting observables, the product rule is satisfied in contrast to PPSs. In addition, for this type of contextuality with PPSs, either the pre-selected or the post-selected state must be the EPR state (with an additional assumption of invariance under exchange of particles).

## 4 PPS AND CONTEXTUALITY IN HIGHER DIMENSIONS

A future article [43] will show how the feat presented in the previous section can be done in higher dimensions, e.g. to the GHZ state [42, 37]. In this case we see that it is not possible to replace the spin operators by ordinary numbers (which is what a noncontextual-HVT attempts to do). However, assignments can be correctly made to each of the  $\hat{\sigma}$ 's if post-selection is utilized. Once again, the limitations to these assignments can be seen by using the structure of TSQM: if we try to measure all of these observables together, then some of the values will be assigned in the diagonal-PPS sense, and therefore measuring these observables will cause a disturbance even though they commute.

Consider the GHZ case of 3 spins pre-selected in the state:

$$|\Psi_{in}\rangle = \frac{1}{\sqrt{2}}|\uparrow_z^1\uparrow_z^2\uparrow_z^3\rangle - \frac{1}{\sqrt{2}}|\downarrow_z^1\downarrow_z^2\downarrow_z^3\rangle \quad (4.61)$$

Consider that the pre-selected state is an eigenstate of the following operators:  $\hat{A}_1 \equiv \hat{\sigma}_x^1\hat{\sigma}_y^2\hat{\sigma}_y^3$ ,  $\hat{A}_2 \equiv \hat{\sigma}_y^1\hat{\sigma}_x^2\hat{\sigma}_y^3$ , and  $\hat{A}_3 \equiv \hat{\sigma}_y^1\hat{\sigma}_y^2\hat{\sigma}_x^3$  with eigenvalue  $+1$ . Also, the pre-selected state is an eigenstate of:

$$\hat{A}_4 \equiv \hat{\sigma}_x^1\hat{\sigma}_x^2\hat{\sigma}_x^3 \quad (4.62)$$

with eigenvalue,  $-1$ , and finally  $\hat{A}_1\hat{A}_2\hat{A}_3 = -\hat{A}_4$ . However, because  $\hat{A}_1$ ,  $\hat{A}_2$ ,  $\hat{A}_3$ , and  $\hat{A}_4$  commute, and because  $\hat{\sigma}_x^i$ ,  $\hat{\sigma}_y^j$ , and  $\hat{\sigma}_z^k$  commute one may ask whether it is possible to also satisfy the above equations by replacing the spin operators by ordinary numbers  $\hat{\sigma}_x^1 = \pm 1$ ,  $\hat{\sigma}_y^1 = \pm 1$ , which is what a non-contextual HVT attempts to do. Without post-selection, this cannot be done because the assignments are, once again, inconsistent with the multiplicative structure of the observables because  $\hat{A}_1\hat{A}_2\hat{A}_3 = 1$ , a contradiction.

### 4.1 GHZ and PPS

Assignments can be correctly made to each of the  $\hat{\sigma}$ 's if post-selection is utilized with limitations again arising from diagonal assignments. E.g., consider a post-selection

of the three particles by measuring their  $x$  component with  $\hat{\sigma}_x = -1$ :

$$\langle \Psi_{fin} | = \langle \downarrow_x^1 \downarrow_x^2 \downarrow_x^3 | \quad (4.63)$$

From the post-selected state we *strongly* know that  $\hat{\sigma}_x^1 \hat{\sigma}_x^2 \hat{\sigma}_x^3 = -1$  and from the pre-selected state we *strongly* know that  $\hat{\sigma}_x^i \hat{\sigma}_y^j \hat{\sigma}_y^k = +1$ . Thus statements of the form  $\hat{\sigma}_y^j \hat{\sigma}_y^k = -1$  can only be made when information is used from both the pre-selected vector *and* from the post-selected vector. For example, if a measurement of  $\hat{\sigma}_y^1 \hat{\sigma}_y^2$  is performed, then we will definitely find  $\hat{\sigma}_y^1 \hat{\sigma}_y^2 = -1$ . However, if we attempt to perform a second measurement, e.g. of  $\hat{\sigma}_y^3 \hat{\sigma}_y^1$  then we will not find  $\hat{\sigma}_y^3 \hat{\sigma}_y^1 = -1$ , because the  $\hat{\sigma}_y^1 \hat{\sigma}_y^2$  measurement will destroy the pre-selected vector which contains the  $\hat{\sigma}_x^i \hat{\sigma}_y^j \hat{\sigma}_y^k = +1$  information that the  $\hat{\sigma}_y^3 \hat{\sigma}_y^1$  measurement depends on. This disturbance occurs even though  $\hat{\sigma}_y^1 \hat{\sigma}_y^2$  and  $\hat{\sigma}_y^3 \hat{\sigma}_y^1$  commute!

## 4.2 Weak Values in GHZ state

We may also consider WMs of the GHZ observables. With the post-selection  $\langle \Psi_{fin} | = \langle \downarrow_x^1 \downarrow_x^2 \downarrow_x^3 |$ , then in the intermediate time we can replace  $\hat{\sigma}_x^1 = \hat{\sigma}_x^2 = \hat{\sigma}_x^3 = -1$  and taking the inner product with this post-selection  $|\Psi_{fin}\rangle$ , we then find:

$$(\hat{\sigma}_y^2 \hat{\sigma}_y^3)_w \equiv \frac{\langle \Psi_{fin} | \hat{\sigma}_y^2 \hat{\sigma}_y^3 | \Psi_{in} \rangle}{\langle \Psi_{fin} | \Psi_{in} \rangle} = -1 \quad (4.64)$$

$$(\hat{\sigma}_y^2 \hat{\sigma}_y^3)_w \equiv \frac{\langle \Psi_{fin} | \hat{\sigma}_y^1 \hat{\sigma}_y^2 | \Psi_{in} \rangle}{\langle \Psi_{fin} | \Psi_{in} \rangle} = -1 \quad (4.65)$$

$$(\hat{\sigma}_y^2 \hat{\sigma}_y^3)_w \equiv \frac{\langle \Psi_{fin} | \hat{\sigma}_y^1 \hat{\sigma}_y^3 | \Psi_{in} \rangle}{\langle \Psi_{fin} | \Psi_{in} \rangle} = -1 \quad (4.66)$$

Using again the analogy with particles in boxes, eq.(4.64) means that particle 2 and particle 3 are not together in the same box, while eq. (4.65) means that particle 1 and particle 2 are not together in the same box, and eq. (4.66) means that particle 1 and particle 3 are not together in the same box. But we only have 2 boxes, so if 1 and 2 are not in the same box and 1 and 3 are not in the same box, then 2 and 3 must be in the same box. It is clear that the above equalities cannot be satisfied simultaneously by replacing the operators for classical numbers taking the values  $\pm 1$ . To simplify this analysis, we define

$$(\hat{\mathbf{N}}_{+++})_w = |\uparrow_y\rangle\langle\uparrow_y|_1 \otimes |\uparrow_y\rangle\langle\uparrow_y|_2 \otimes |\uparrow_y\rangle\langle\uparrow_y|_3, \quad (4.67)$$

where the 2 boxes are denoted by  $\pm$  referring to the spin component along  $y$ . Using Theorems 1 and 2, it can be shown [43] that:

$$(\hat{\mathbf{N}}_{+++})_w = \hat{\mathbf{N}}_{---} = -\frac{1}{4} \quad (4.68)$$

and

$$(\hat{\mathbf{N}}_{++-})_w = (\hat{\mathbf{N}}_{--+})_w = (\hat{\mathbf{N}}_{+-+})_w = (\hat{\mathbf{N}}_{-+-})_w = \dots = \frac{1}{4} \quad (4.69)$$

We have thus shown how to obtain a non-classical negative triplet WVs, which again, cannot be reproduced by a noncontextual-HVT.

## 5 CONCLUSION

Mermin's results reviewed in §'s 3.2.1-3.2.4 were characterized by Mermin as "...what follows is not idle theorizing about 'hidden variables'. It is a rock solid quantum mechanical effort to answer a perfectly legitimate quantum mechanical question." [27] BKS showed that noncontextual-HVT's are not possible in general. Therefore, an interesting approach to HVT's and BKS is whether anything new can be learned about experimental situations, in the same spirit as Shimony's apt phrase "experimental metaphysics." For example, Bell's Theorem led to interesting experiments which tested the notion of whether quantum entanglement could be stronger than classical correlations. Another example is Hardy's-paradox [45] (HVTs for position could not be assigned) which was traditionally "resolved" by arguing that measurements to verify the paradox could not be implemented simultaneously and therefore Hardy's-paradox was purely a formal result without empirical consequences. However in [11, 30], we demonstrated that WMs could be implemented simultaneously on the paradoxical statements (with experimental results [36]). E.g. the electron is always on the overlapping path ( $N_{Ow}^- = 1$ ) and the positron is always on the overlapping path ( $N_{Ow}^+ = 1$ ), yet they are never there together ( $N_{O,Ow}^{+,-} = 0$ ). Again, non-classical WV's were demonstrated [11] by the negative pair-wise occupation,  $N_{NO,NOw}^{+,-} = -1$ . This again was a violation of the product rule 1.3 ( $N_{Ow}^- N_{Ow}^+ \neq N_{O,Ow}^{+,-}$ ) and a manifestation of diagonal-PPSs and non-classical WV's. Nevertheless, despite their non-classical behavior, WV's do obey a simple, intuitive, and, most important, *self-consistent* logic.

In this paper, we have carried this and the program initiated by Mermin further and have shown new ways that the "charming elementary mathematics" of BKS can manifest empirically. E.g. in the 3-box-paradox/diagonal-PPS of §2,  $\hat{\mathbf{P}}_A = 1$  if only box  $A$  is opened, while  $\hat{\mathbf{P}}_B = 1$  if only box  $B$  is opened, but if we measure both box  $A$  and box  $B$ , then the particle will not be found in both boxes, i.e.  $\hat{\mathbf{P}}_A \hat{\mathbf{P}}_B = 0$  even though  $\hat{\mathbf{P}}_A$  and  $\hat{\mathbf{P}}_B$  commute, a violation of the product rule. If WMs are performed, then the non-classical result  $(\hat{\mathbf{P}}_C)_w = -1$  is obtained. In the Mermin case §3.2, the violation of the product rule and diagonal-PPS is that  $\hat{\sigma}_x^1 \hat{\sigma}_y^2 \hat{\sigma}_x^2 \hat{\sigma}_y^1 = -1$  (given the PPS of fig. 4.a) even though separately  $\hat{\sigma}_x^1 \hat{\sigma}_y^2 = +1$  and  $\hat{\sigma}_x^2 \hat{\sigma}_y^1 = +1$  and non-classical WV's are again obtained,  $(\hat{\mathbf{N}}_{--})_w = -\frac{1}{2}$ . Similar results were obtained in the 4-D/EPR case (i.e.  $\hat{\sigma}_x^1 \hat{\sigma}_y^2 = -1$  even though  $\hat{\sigma}_x^1 = -1$  and  $\hat{\sigma}_y^2 = -1$ ) and also in the 8-D GHZ case.

How general are these considerations? Consider any set of commuting observables  $\hat{A}_1, \hat{A}_2, \hat{A}_3$  and any other set of commuting observables  $\hat{B}_1, \hat{B}_2, \hat{B}_3$ . Suppose we perform pre-selection measurements on  $\hat{A}_1, \hat{A}_2, \hat{A}_3$  and perform post-selection measurements on  $\hat{B}_1, \hat{B}_2, \hat{B}_3$ , and suppose, e.g.  $[\hat{A}_2, \hat{B}_2] \neq 0$ . Products of observables, e.g.  $\hat{A}_1 \hat{A}_2$ , can be ascertained solely based on the pre-selection, while products of observables, e.g.  $\hat{B}_2 \hat{B}_3$ , can be ascertained solely based on the post-selection. However, as soon as we change the sequence of measurements of  $\hat{A}_2$  and  $\hat{B}_2$ , then it is a diagonal-assignment and they will interfere with each other in PPS. Given this, it is evident that there are versions of BKS for which we cannot obtain definite results for IMs utilizing PPSs. It is not the case that everything that cannot be known by one vector can be known by two vectors: there are many sets of operators for which it is impossible to assign all the possible eigenvalues via PPS. However, our arguments

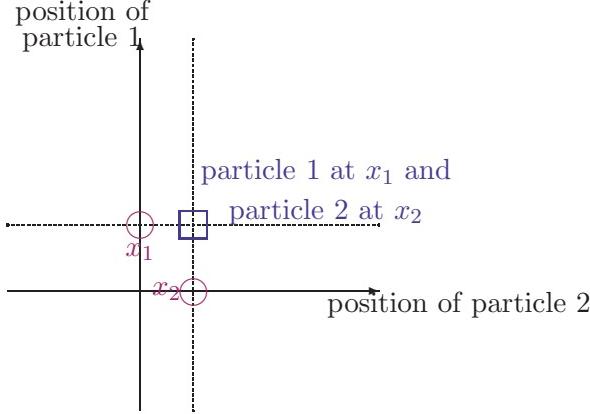


Figure 9: Weak Value density for 2 particles

concerning WMs are completely general.

The results of this article could be used to further exploit the differences between classical and quantum information: while the study of the non-classical aspects of entanglement (Einstein-Podolsky-Rosen/Bohm) started as a foundational examination of HVTs, it was subsequently probed experimentally and used as a resource for quantum computation and communication. Similarly, using TSQM [12], Englert demonstrated a new form of cryptography [19] and experiments based on the optical version of this problem [18] were successfully performed. With respect to contextuality, BKS is often based solely on the structure of *operators*, while quantum information applications require special *states*. In this article, we have shown how to probe and empirically manifest contextuality through PPS *states*, thereby opening new possibilities of utilizing contextuality in these applications.

We end with a conjecture: if we start with a non-classical situation as reflected in BKS or a violation of Bell's-inequality, then we can always find a post-selection which can empirically manifest non-classical WV. This thereby holds the possibility of shifting foundational debates into a new issue because we can now probe non-classicality in actual WM experiments. Non-classical WV follow intuitively from the logic of WV: e.g. consider a generic multi-particle systems in a higher dimensional space  $\rho(x_1, x_2) \sim \Psi_{fin}^*(x_1, x_2)\Psi_{in}(x_1, x_2)$  and consider also projections onto space and time. In general,  $\rho(x_1, x_2)$  cannot be measured locally because there is no local way of measuring particle 1 at  $x_1$  simultaneously with measuring particle 2 at  $x_2$ : this would involve a nonlocal Hamiltonian. However, the projection on each line can be measured separately: if it is known that particle 2 is at  $x_2$  then this means  $\int \rho(x_1, x_2) dx_1 = 1$ . Similarly, if it is known that particle 1 is at  $x_1$  then this means  $\int \rho(x_1, x_2) dx_2 = 1$ . However, asking the question if both particles are there together is another point in phase space (see fig. 9). The particle also could have been located simultaneously in another position with certainty, but then we need to place a negative number somewhere else in order to satisfy the global constraint, i.e. the integral of the WV has to add up to 1 (because there is just a single particle along each line), but the individual numbers at each point can be arbitrary. IM outcomes reveal an integration along just one line, but the WV densities are not just those lines. E.g. in the Hardy case the seeming contradiction that both particles are there

individually but are not there together is resolved by a negative number of particles at another point in phase space.

While there is an operational or experimental meaning to a density over a set of commuting observables, as shown in §2.2.1, there is no experimental meaning for a density over a set of non-commuting observables even though such densities may have formal utility as an aide to calculation [46]. As a result of this dis-connect between densities over non-commuting observables and IMs, there is a similar problem with **WMs and WV**s. An attempt to give an operational or experimental meaning to WMs of densities over non-commuting observables will illustrate the same problem as demonstrated in 2.2.1, even though in this article we demonstrated the ability to define WMs and WV for a **discrete** set of non-commuting observables.

**Acknowledgments:** JT thanks Yakir Aharonov for many conversations and the Templeton foundation for support.

## References

- [1] Y. Aharonov, D. Bohm. Phys. Rev. **122**, 1649 (1961); reprinted in *Quantum Theory and Measurement*, eds. J. A. Wheeler and W. H. Zurek (Princeton University Press), 1983; Y. Aharonov, H. Pendleton, and A. Petersen, International J. Theor. Physics, Vol. 2, 213(1969).
- [2] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz, Phys. Rev. **134**, B1410 (1964), reprinted in *Quantum Theory and Measurement*, eds. J. A. Wheeler and W. H. Zurek (Princeton University Press), 1983, pp. 680-686.
- [3] D. Albert, Y. Aharonov, and S. D'Amato, *Phys. Rev. Lett.* **54**, 5 (1985).
- [4] L. Vaidman, Y. Aharonov and D. Albert, *Phys. Rev. Lett.* **58**, 1385 (1987).
- [5] Y. Aharonov, L. Vaidman, Phys. Rev. A, **41**, (1990), 11.
- [6] Y. Aharonov and L. Vaidman, J. Phys. A **24**, 2315 (1991).
- [7] Y. Aharonov and L. Vaidman, *Phys. Lett. A* **178**, 38 (1993).
- [8] Y. Aharonov, L. Davidovich, N. Zagury, *Phys. Rev. A* **48** (1993) 1687.
- [9] B Reznik, Y. Aharonov, Phys. Rev. A, **52**, p. 2538, (1995).
- [10] Y. Aharonov, S. Massar, S. Popescu, J. Tollaksen, and L. Vaidman, Phys. Rev. Lett., **77**, p. 983, (1996).
- [11] Aharonov Y, Botero A, Popescu S, Reznik B, Tollaksen J, PHYS LETT A 301 (3-4): 130-138 AUG 26 2002.
- [12] Yakir Aharonov and B-G. Englert, , Z. Naturforsch A 56, 16-19, (2001).
- [13] Aharonov, Yakir, Rohrlich, Daniel, *Quantum paradoxes: quantum theory for the perplexed*, Weinheim ; Cambridge : Wiley-VCH, 2005.

- [14] Aharonov, Y., Botero, A., Phys. Rev. A 72, 052111 (2005).
- [15] A. Botero and B. Reznik, Phys. Rev. A, Vol. 61, 050301-4, (2000).
- [16] Bub, J., Phys. Rev. A 63 (2001) 032309.
- [17] Aharonov, Y., and Tollaksen, J, forthcoming.
- [18] B.-G Englert, C. Kurtsiefer, H. Weinfurter Phys Rev A, Vol. 63, art. 032303.
- [19] A. Beige, B-G. Englert, C. Kurtsiefer, and H. Weinfurter, quant-ph/0101066.
- [20] Bell , Physics, 1964.
- [21] Bell, John S. (1966): Reviews of Modern Physics 38: 447-52; reprinted in Bell, John S. (1987): Speakable and Unspeakable in Quantum mechanics (Cambridge; Cambridge University Press).
- [22] Chiao RY, Steinberg AM, Prog in Opt 37: 345-405 1997; Solli DR, McCormick CF, Chiao RY, et al. Phys Rev Lett 92 (4): Art. No. 043601 JAN 30 2004.
- [23] M. Gell-Mann and J.B. Hartle, Phys. Rev. **D 47**, 3345 (1993).
- [24] Kochen, Simon and Specker, Ernst (1967): Journal of Mathematics and Mechanics 17: 59-87.
- [25] M. S. Leifer, R. W. Spekkens, Int. J. Theor. Phys. 44:1977-1987 (2005); quant-ph/0412179.
- [26] M. S. Leifer, R. W. Spekkens, Phys. Rev. Lett. 95, 200405 (2005).
- [27] N. David Mermin, Phys. Rev. Lett., **74**, p. 831,(1995); R.S. Cohen, et. al. (eds), *Potentiality, Entanglement and passion-at-a-Distance*, 149-157,Kluwer, 1997.
- [28] N. David Mermin, Rev. Mod. Phys., **65**, p. 803,(1993).
- [29] L. Vaidman, *Phys. Rev. Lett.* **70**, 3369(1993). in *Symposium on the Foundations of Modern Physics*, 3, P.J. Lahti, P. Bush, and P. Mittelstaedt (eds.), pp. 406-417, (World Scientific,Cologne 1993).
- [30] Jeff Tollaksen, 2001 PhD thesis, Boston University, thesis advisor: Yakir Aharonov.
- [31] Tollaksen, J. “Robust Weak Measurements,” forthcoming; and Y. Aharonov, J. Tollaksen, “Non-statistical Weak Measurements,” forthcoming.
- [32] Aharonov Y., Gruss, E., quant-ph/0507269.
- [33] Gleason, A., J. Math. Mech., **6** 885 (1957).
- [34] A. M. Steinberg, *Phys. Rev. Lett.* **74**, 2405(1995).

- [35] Resch KJ, Lundeen JS, Steinberg AM PHYSICS LETTERS A 324 (2-3): 125-131 APR 12 2004.
- [36] Lundeen JS, Resch KJ, Steinberg AM PHYSICAL REVIEW A 72 (1): Art. No. 016101 Part A-B JUL 2005; S.E. Ahnert and M.C. Payne, Phys. Rev. A 70, 042102 (2004)
- [37] L. Vaidman, Found. Phys., **26**, (1996), 895.
- [38] J. von Neumann, *Mathematical Foundations of Quantum Theory*, Princeton, University Press, New Jersey (1983).
- [39] Simon, C., Zukowski, M., Weinfurter, H., Zeilinger, A. (2000), Phys. Rev. Lett., 85, 1783-1786.
- [40] Kastner, R., Found. Phys 29 (6), 851.
- [41] Kirkpatrick, K., Journal of Physics A, 36, 4891.
- [42] D.M.Greenberger, M.A.Horne and A.Zeilinger, in Bell's Theorem, Quantum Theory and Conceptions of the Universe, M.Kafatos, Ed. (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1989) p.69; see also D.M.Greenberger, M.A.Horne, A.Shimony and A.Zeilinger, Am.J.Phys. 58, 1131-43 (1990).
- [43] Aharonov, et al, “Families of weak values,” forthcoming.
- [44] T Ravon, L Vaidman, “The 3-box paradox re-visited,” quant-ph/0606067.
- [45] L. Hardy, Phys. Rev. Lett. **68**,(1992), 2981.
- [46] Aharonov, Y., Tollaksen, J. “Quantum Hidden Variables,” forthcoming.
- [47] J. Bub and H.Brown, *Phys. Rev. Lett.*, 56: 2337, 1986.
- [48] Ognyan Oreshkov, Todd A. Brun, Phys. Rev. Lett. 95, 110409 (2005).
- [49] A. Peres, J. Phys. A 24, L175 (1991).
- [50] Feynman, R. “Negative Probability,” in *Quantum Implications* B.J. Hiley and F.D. Peat (eds.) . New York: Routledge & Kegan Paul, pp.224-226.